Hand Gesture Recognition based on Fusion of Moments

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Technology

In

Signal and Image Processing

By

SUBHAMOY CHATTERJEE

Roll No: 212EC6185

Department of Electronics and Communication Engineering

National Institute Of Technology, Rourkela

Orissa 769 008, INDIA

2014

Hand Gesture Recognition based on Fusion of Moments

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Technology

In

Signal and Image Processing

By

SUBHAMOY CHATTERJEE

Roll No: 212EC6185

Under the Guidance of

Dr. Samit Ari

Assistant Professor

Department of Electronics and Communication Engineering

National Institute Of Technology, Rourkela

Orissa 769 008, INDIA

2014

Dedicated to

My beloved mother Mrs. Barnali Chatterjee

My respected father Mr. Saktimoy Chatterjee

My dear sister Ms. Tithi Chatterjee

My friend Mr. Manu Thomas

And all my well wishers

NATIONAL INSTITUTE OF TECHNOLOGY
ROURKELA

CERTIFICATE

This is to certify that the thesis titled **"Static Hand Gesture Recognition based on Fusion of Moments"** submitted by Mr. **Subhamoy Chatterjee** in partial fulfilment of the requirements for the award of Master of Technology degree in Electronics and Communication Engineering with specialization in "Signal and Image Processing" during session 2012-2014 at National Institute Of Technology, Rourkela is an authentic work by his under my supervision and guidance. To the best of my knowledge, the matter embodied in the thesis has not been submitted to any other university / institute for the award of any Degree or Diploma.

Date:

Dr. Samit Ari
Assistant Professor
Dept. of Electronics and Comm. Engineering
National Institute of Technology
Rourkela-769008

DECLARATION

I hereby declare that the work presented in the thesis entitled as "***Hand Gesture Recognition based on Fusion of Moments***" is a bona fide record of the systematic research work done by me under the guidance of **Prof. Samit Ari**, Department of Electronics & Communication, National Institute of Technology, Rourkela, India and that no part thereof has been presented for the award of any other degree.

Subhamoy Chatterjee

(Roll no. 212Ec6185)

Acknowledgement

I would like to thank my supervisor **Prof. Samit Ari** for his guidance, advice and constant support throughout my thesis work here in National Institute of Technology, Rourkela.

I am highly grateful to all the faculty members and staff of the Department of Electronics and Communication Engineering, N.I.T. Rourkela for their unforgotten teaching and motivation in my research work.

I would like to thank all my friends, lab mates and especially my two dear friends Manu Thomas and Abhinav Kartik for giving me their valuable advises and ideas, which grew a greedy interest for me towards my research area. I would like to express my gratitude to Mr. Manu Thomas who has become an idol for all his lab mates including me.

I would like to show my gratitude for my parents, for their sacrifice throughout my life. They are the inspiration of my research work.

Date:

Time:

Subhamoy Chatterjee

212EC6185

Signal and Image Processing

ABSTARCT

Static hand gesture recognition can be applied in various domains such as human-computer interaction (HCI), remote control, robot control, virtual reality etc. Hand gesture recognition is mainly the study of detection and recognition of various hand gestures like American Sign Language hand gestures, Danish Sign Language hand gestures etc by a computer. This work is focussed on three main issues in developing a gesture recognition system. These are (i) Threshold independent skin colour segmentation using Modified K-means clustering and Mahalanobish distance (ii) illumination normalization (iii) user independent gesture recognition based on fusion of Moments. A vision based static hand gesture recognition algorithm which consists of three stages: pre-processing, feature extraction and classification, is presented in this work. It is very challenging to segment hand regions from the static hand gesture colour images, due to varying light conditions and complex background. Since skin pixels can vary with different illumination condition, to find the range of skin pixels, becomes a hard task in case of colour space based skin colour segmentation. This work proposes a semi-supervised learning algorithm based on modified K-means clustering and Mahalanobis distance to extract human skin colour regions from the static hand gesture colour images. An efficient illumination invariant algorithm based on power law transform and averaging RGB colour space is proposed. Normalized binary silhouette is extracted from the hand gesture image and background and object noise is removed by Morphological filtering. Non-orthogonal moments like geometric moments and orthogonal moments like Tchebichef and Krawtchouk moments are used here as features. The Krawtchouk moment features are found to be very effective in hand gesture recognition compared to Tchebichef and Geometric moment features. To make the system real time efficient, different users are used for training and testing. In user-independent situation, neither of these moments has shown efficient classification accuracy. To improve the performance of classification, two feature fusion strategies have been proposed in this work; serial feature fusion and parallel feature fusion. A feed-forward multi-layer perceptron (MLP) based artificial neural network classifier is used in this work as a classifier. The proposed two fusion based moment features especially parallel fusion of Krawtchouk and Tchebichef moment has shown better performance as user-independent. The proposed hand gesture recognition system can be well realized for real time implementation of gesture based applications.

List of Figures

List of Tables

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 HAND GESTURE RECOGNITION SYSTEM

We mainly communicate with others with the help of our voice and body language. Although speech is the mostly used way of interaction for human beings, body language and facial expressions are also used to interact with others. In many cases, interaction with the physical world by body language and gestures is more reliable. Gestures or body languages can be expressed in various ways. These can be expressed by simply waving hands, making a meaningful gesture by hand, finger or body pose or by a meaningful facial expression. In between these gestures and expressions, hand gestures are the most efficient means to express meaningful and significant information. In our real life situation, we use hand gestures to communicate with mute and deaf people by using sign languages, to count numbers and to express a feeling like 'good bye' or 'stop'. With the recent developments in Artificial intelligent, Soft Computing and Neural Networks, hand gestures are becoming the most important tool to interact with computers and machines. Now a days, gesture based computer control is one of the developing research field in Pattern Recognition. Even many industries have also started to implement Human Computer Interaction techniques to make machines more intelligent.

Generally, gestures are of two types: Static gestures and Dynamic gestures. Static gestures are mainly expressed by some meaningful body expressions. Simple 'Stop' gesture is a static hand gesture to express a significant information. Many times static gestures are used to express dynamic gestures also. In American Sign Language, digits greater than nine have been expressed by the hand movement and with a static gesture from zero to nine. Dynamic gestures are expressed by body movements. Simply waving hand to express 'good bye' is a dynamic gesture. Compared to other body parts, hands are more flexible. For that reason, hand gestures are mostly used in human computer interaction than other body parts.

The development of static and dynamic hand gesture recognition system depends solely on the image acquisition and processing technology. With the recent developments in image acquisition technology and with the invention of highly reliable cameras, both static and dynamic gestures are becoming the most important tool for human computer interaction. Hand gestures are on the way to replace the commonly used input devices like mouse, keyboard, joystics and some special pens. Even it is thought that hand gestures will replace the touch screen technology of mobile and other devices very soon. Many companies have

started to develop gesture based computer control technologies, but a lot development in this field is needed.

Some researchers have employed gloves and similar type of hardware for static hand gesture recognition [1] using some expensive sensors. So, these methods are very complicated in real time applications. For that reason, vision based static hand gesture recognition techniques are mostly used in real time applications. These methods need no hardware except a camera, so these methods are quite cheaper and can be easily accepted by any industry.

1.2 GESTURES

We are interested in recognizing American Sign Language (ASL) [2] human hand gestures using computer vision principles. Gestures are usually understood as hand and body movement that can convey information and can be a proper means of communication between two persons. According to Webster's dictionary:

A gesture is a pose or movement of some organs like hand, finger etc to convey certain kind of meaningful information. It is one of the important medium to communicate with others. Gestures are divided into two categories: I) Static gesture [3] II) Dynamic gesture [4].

A dynamic gesture is just movement of hand or any other body part over a period of time whereas a static gesture is a pose or position of hand or any other body part. Example of dynamic gesture may be just waving a goodbye and example of static gesture may be the stop sign. Some complex algorithms and methods are designed to understand and interpret various types of gestures over a period of time. This complex processes are called gesture recognition.

1.3 GESTURE BASED APPLICATIONS

Static and dynamic hand gestures have a huge application in both multidirectional control and sign language purpose.

Robot control in inaccessible remote areas: It is impossible for human beings to operate and physically present in a hostile condition like nuclear power plant, defence research plants,

medicine manufacturing plants etc. In that case, a robot and gesture based robot control systems have a great use. Often it is impossible for human operators to be physically present near the machines [5]. Some technical assistance and knowledge can be provided to the robot system through gesture recognition systems. Recently researchers of University of California, San Diego have designed real time ROBOGEST system [6] which aims to control an outdoor autonomous vehicle by sign language recognition system of hand gestures .

Virtual reality: It is a computer aided environment which is analogous to real world. It is designed by a high quality of animation technology. It can be displayed through computer screen as well as special stereoscopic displays. Many companies have focussed its research in virtual reality by hand gesture control because of its prosperous application in medical and gaming technology.

Sign Language: In sign languages with some movements and poses of body parts we communicate with mute and deaf people. It can be expressed by hand poses, hand gestures, movement of hands etc. It has a huge similarity with speaking languages, that's why these languages are called natural languages. Sign languages are mainly used to communicate with deaf and dumb people, in sports, for religious practices and even in our daily life as well as workplace [2]. Sign language recognition is one of the mostly used human computer interaction (HCI) application [7].

Remote control Static hand gesture recognition systems are on the run to replace the remote control devices of electronic gadgets. It is more reliable to control electronic gadgets by meaningful hand gestures than remote control devices. By applying a proper gesture recognition algorithm remote control [8] with the wave of a hand or pose of a hand of various devices is possible. Researchers have designed a proto-type system called WiSee [9] which is capable for remote controlling of electronic gadgets. The proposed system is connected using Wi-Fi and uses gestures such as waving our arms, punching, and kicking. This system can be used in all of our daily life applications such as turn out lights, control the television, music system, or even a room's thermostat. The program can take commands from up to 5 users and is understood to not be triggered by the usual movements of people in the house.

Automobile control: Automobile industry is also in the run to use gesture based control of automobiles and its various accessories. Researchers are designing some gesture recognition systems for blind-spot recognition and parking assist. Some researchers are designing some automobiles which can be driven by using some gestures. These types of automobiles can be

used in adverse atmospheres where human beings are not safe. HUDs (Head-up Displays) and Garmin have designed some gesture controlled cars [10] which can be used in diverse condition as well as for luxury purpose.

Affective computing: With the increasing development of Artificial Intelligence, it is now possible for computers to recognize and analyze human sentiment and imotions through a vision based gesture recognition algorithm. This field of study is called affective computing. By using one or more cameras we can analyze automatic emotion detection of a human beings.

A simple gesture which is used universally by waving or pose of one hand means either 'hi' or 'goodbye'. We can communicate to persons with different languages without knowledge of their language using hand gestures and sign languages. Various sign languages are the only means of communication for mute and deaf persons.

American [2] and Danish sign language [11] have representation for words. So by a sign language recognition system we can recognize a speech or a hidden message also. Sign language recognition system can also be used as an interpreter between mute and unsigned persons.

Although hand gesture recognition system has a huge application in all the above mentioned fields, it needs more developments for real time applications. Here in this work we have tried to implement a real time efficient static hand gesture recognition system.

.

1.4 LITERATURE SURVEY

There are mainly 3 primary issues in a hand gesture recognition system: (i) hand detection and separation of region of interest area from the captured image, (ii) Illumination, rotation and scale normalization (iii) User-independent gesture recognition. Ong and Ranganath [12] described a detail analysis of hand gesture recognition system and the difficulties to implement a hand gesture recognition system.

At first hand gesture recognition algorithms were implemented in a uniform background. A number of restrictions were imposed on the work. Because of this uniform background, segmentation was comparatively easy. In uniform background, skin color detection based on

both color space models [13] and clustering techniques [14] could be easily implemented. In case of non-uniform background, extraction of binary hand silhouette from the image is the most challenging work in hand gesture recognition algorithm. Some segmentation works in a non-uniform background has been described in [3, 13–14]. To make the algorithm invariant in size, orientation and illumination some normalization techniques have been proposed in [13–14].

Mainly two types of features are used in hand gesture recognition: a) shape based features, b) contour based features. Contour based features are not related to the shapes of the image, they are calculated based on the boundary profiles of the image. On the other hand, shape based features are calculated on every pixel of the image. A well-known shape based feature, Orientation of histogram has been proposed and analyzed in [16, 17]. These features have shown significant result in classification and are known as illumination invariant features. Amin and Yan [18] have used Gabor filter features which were computed from raw color image of the hand gestures. For dimensionality reduction of the feature sets the principal component analysis (PCA) has been proposed. In [19] Moments have been proposed as feature. Some features are directly calculated from the hand images like number of finger, distance and angles between the fingers etc [20]. A real time efficient static hand gesture recognition system using Krawtchouk moment as feature has been proposed in [3]. Researchers have also proposed a rotation and scale invariant gesture recognition algorithm in this paper. In user independent situation this method does not show satisfactory classification accuracy, because of different hand shapes in testing and training database.

From the above mentioned analysis, it can be inferred that Contour based techniques do not perform well for gestures with almost analogous size. But in case of user independent gesture classification, shape based features do not work well. So in user-independent gesture classification, some feature level or decision level fusion techniques should be employed to reduce the misclassification in gesture recognition.

1.5 SYSTEM OVERVIEW

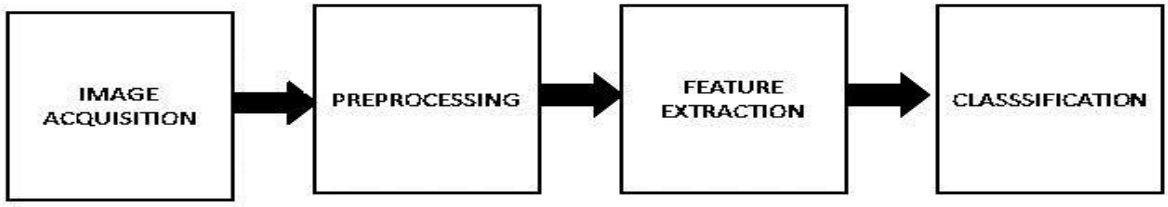


Figure 1. 1: System Overview

Vision based static hand gesture recognition algorithm is tantamount to human perception about their surroundings and it is very difficult to implement. Many researchers have proposed different ideas and methods for vision based static hand gesture recognition system.

Some researchers have used three dimensional model of hand for template matching [3]. Using a well-known classifier they have classified hand gestures. This method is quite complicated compared to camera based methods.

There is an alternative method for vision based gesture recognition. In this method some gesture images are captured by one or more cameras. Then the whole database is split into training and testing. Using some appropriate feature extraction techniques testing and training features are extracted. Depending on the training feature set, classifier is trained and testing feature set is used to test the network [3].

In this work, we have used camera based method for gesture recognition. We have made two hand gesture databases. In case of first database, a uniform black background is used behind the user to avoid background noises. Second one is taken in complex background. In both the databases, forearm region is separated by wrapping a black cloth. In second database hand region is restricted to have maximum area compared to other regions.

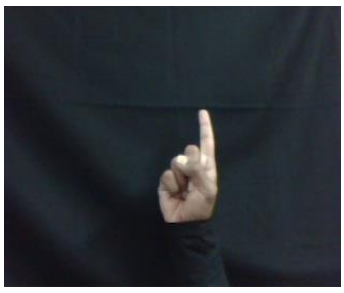
Segmentation, morphological filtering and some rotation and illumination normalization techniques are applied on images in pre-processing phase. Then using orthogonal (Krawtchouk and Tchebichef moment) and non-orthogonal (Geometric moment) moment features we have calculated training and testing feature sets. In user-independent case, none of these moments has shown satisfactory classification accuracy. To increase classification accuracy in user-independent condition two feature level fusion strategies have been proposed: a) serial feature fusion b) parallel feature fusion. We have used Artificial Neural Network classifier to classify hand gesture images.

1.6 DATABSE DESCRIPTION

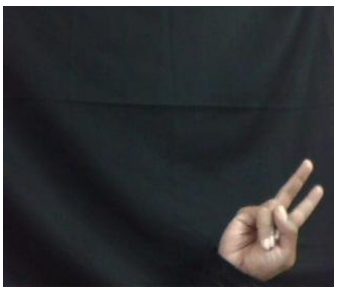
In this project all operations are performed on RGB colour images .We have made two hand gesture databases. In case of first database, a uniform black background is used behind the user to avoid background noises. Second one is taken in complex background. In both the databases, forearm region is separated by wrapping a black cloth. In second database hand region is restricted to have maximum area compared to other regions.

A Logitech c120 webcam has been used to capture the hand gesture images. The resolution of grabbed image is 320 × 240 for both the databases. All images are taken in various angles and in different light conditions to make our gesture recognition algorithm rotation and illumination invariant. We have used the uniform background database only for semi-supervised learning purpose. Second database is mainly used for testing and training purpose. The dataset consists of 1500 gestures of 10 classes, 15 samples each class of 10 users. The dataset is equally divided into training and testing datasets of 750 gestures of 10 classes for 5 different users to make the system user-independent.

Two databases are given below



ASL Digit 1



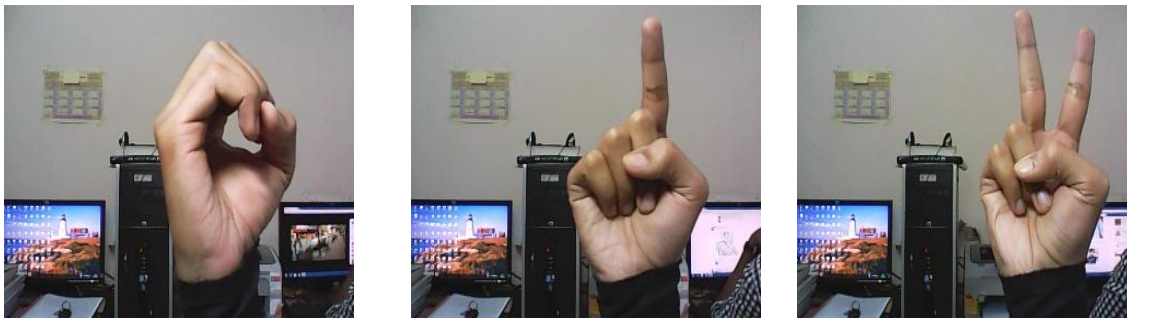
ASL Digit 2

ASL Digit 3

ASL Digit 4 ASL Digit 5 ASL Digit 6

ASL Digit 7 ASL Digit 8 ASL Digit 9

Figure 1. 2:Sample image from Database 1



ASL Digit 0 ASL Digit 1 ASL Digit 2

ASL Digit 3 ASL Digit 4 ASL Digit 5

ASL Digit 6 ASL Digit 7 ASL Digit 8

ASL Digit 9

Figure 1. 3:Sample images from Database 2

1.7 THESIS OUTLINE

In **Chapter 2** Pre-processing of gesture recognition system is described. Pre-processing stage consists of image acquisition, segmentation, rotation normalization, illumination normalization and morphological filtering methods. At first we have segmented our hand gestures using YCbCr skin colour segmentation. But this method is not robust in varying illumination condition, and the threshold values for segmentation change with illumination level. To overcome this, we have implemented a segmentation process by semi-supervised learning algorithm for skin colour segmentation based on K-means clustering and Mahalanobis distance. Morphological operations have been used to get the original shape of the binary hand silhouette and to remove object noise. Some algorithms have been discussed to make the system rotation and illumination invariant.

In **Chapter 3** we have extracted features from the binary silhouette. Here orthogonal moments namely Krawtchouk and Tchebichef moments and non-orthogonal moment namely geometric moment are used as features. To improve classification accuracy in user-independent condition we have proposed two feature fusion strategies: Serial feature fusion and Parallel feature fusion. These two techniques are also discussed here.

We explained classification technique using Artificial Neural Network (ANN) classifier. We have used four parameters to justify classification performance. These are: Accuracy, Sensitivity, Specificity, and Positive Predictivity.

In **chapter 4** we have concluded our work and discussed about its future scope.

References

1. P. Kumar, J. Verma and S. Prasad, "Hand Data Glove: A Wearable real-time Device For Human-Computer Interaction" *International Journal Of Advanced Science And Technology, vol. 43*, Jun. 2012.
2. https://www.nidcd.nih.gov/health/hearing/pages/asl.aspx
3. S. P. Priyal and P. K. Bora "A Robust Static Hand Gesture Recognition System Using Geometry based Normalization and Krawtchouk Moments" *Pattern Recognition, vol. 46, no. 8, pp. 2202-2219,* 2013
4. H. Suk,, S. Bong, and L. Seong. "Hand gesture recognition based on dynamic Bayesian network framework." *Pattern Recognition* vol. 43, no.9, pp. 3059-3072, 2010.
5. M. Manigandan and I. M. Jackin "Wireless Vision based Mobile Robot control using Hand Gesture Recognition through Perceptual Color Space", *International Conference on Advances in Computer Engineering,* Bangalore, India, Jun. 2012.
6. http://globalcatalog.com/robogestsrl.it
7. A.Agrawal, R. Raj, and S. Porwal " Vision-based multimodal human-computer interaction using hand and head gestures", *Conference on Information and Communication Technologies (ICT 2013)*, Tamil Nadu, India, Apr. 2013.
8. U. V. Solanki and N. H. Desai, "Hand gesture based remote control for home appliances : Handmote" *World Congress on Information and Communication Technologies (WICT),* Mumbai, India, 2011.
9. http://wisee.cs.washington.edu/
10. http://www.cs.bham.ac.uk/~rjh/courses/IntroductionToHCI/201314/GroupSubmissions/Group21.pdf
11. http://www.ethnologue.com/language/dsl
12. Ong and Ranganath, "Automatic Sign Language Analysis: A Survey and the Future Beyond Lexical Meaning*," IEEE Trans. Pattern Anal. Mach. Intell,* vol. 27, no. 6, pp. 873-891, June 2005.
13. S. K. Singh, D. S. Chauhan, M. Vatsa and R. Singh " A Robust Skin Color Based Face Detection Algorithm" *Tamkang journal of science and engineering, vol. 6,no. 4, pp. 227- 234, 2003.*

14. R. Vijayanandh and G. Balakrishnan, "Performance Analysis of Human Skin Region Detection Techniques with Face Detection Application", *International Journal of Modeling and Optimization, vol. 1, no. 3*, Aug. 2011.
15. D. K. Ghosh and S. Ari "A Static Hand Gesture Recognition Algorithm Using K-Mean Based Radial Basis Function Neural Network" *In: 8th International Conference on Infor-mation, Communications and Signal Processing (ICICS),* pp. 1-5, Singapore , 2011.
16. T.William, T. Freeman and M. Roth " Orientation histogram for hand gesture recognition" *in proceedings of the 1st International workshop on Automatic face and gesture recognition, pp. 296- 301,* 1995.
17. H. zhou, D. J. Lin and T.S. Huang " Static hand gesture recognition based on local orientation Histogram feature distribution model *", In proceedings of the Conference on Computer Vision and Pattern Recognition workshops, vol. 10, pp. 161*, 2005.
18. M. Amin and H. Yan "Sign language finger alphabet recognition from gabor-pca representation of hand gestures" *in: Proceedings of the International Conference on Machine Learning and Cybermatics, vol. 4, pp. 2218- 2223*, 2007.
19. S. P. Priyal and P. K. Bora "A Study Of Static Hand Gesture Recognition using Moments" *International Conference on Signal Processing and Communications (SPCOM), pp. 1-5*, IISC, Bangalore (2010).
20. S. Chandran and A. Sawa " Real time detection and understanding of isolated protruded fingers", *In Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop, vol. 10, pp. 152*, 2005.

CHAPTER 2

PREPROCESSING OF HAND GESTURE IMAGE

2.1 INTRODUCTION

In Static hand gesture recognition, pre-processing is the primary and the most important step. In pre-processing, binary silhouette of the hand gesture image is extracted for shape based feature extraction. For contour based feature extraction, boundaries are extracted from the colour hand gesture image. In this work, we have used shape based feature extraction techniques, so we have extracted binary hand silhouette as region of interest. Pre-processing consists of 3 steps

(a) Segmentation
(b) Rotation and illumination normalization
(c) Morphological filtering.

In segmentation process, skin colour region is detected and extracted from the captured hand gesture image. Skin colour contains relatively concentrated information in any gesture image. The process of extracting the region of interest from the hand gesture image is called skin colour segmentation. We have employed two different techniques for skin colour segmentation: (i) YCbCr based skin colour segmentation [1] (ii) skin colour segmentation using semi-supervised learning based on modified K-means clustering [2] and Mahalanobish distance. In YCbCr skin colour segmentation threshold values of skin region has been proposed for Asian and European skin colour [3]. Colour-space base skin colour segmentation methods are not robust for skin colour detection, because in varying illumination condition and in complex background, threshold values for the colour space models also vary.

We have proposed a skin colour detection process based on semi-supervised learning, which has shown robustness in varying illumination condition and complex background.

Segmented hand gesture image should be rotation [4] and illumination invariant [5], otherwise gestures of same classes will be misclassified as gesture of different class. For that purpose, we have proposed a rotation normalization and illumination normalization algorithm to make the images rotation and illumination invariant.

Some morphological operations called dilation, erosion, opening and closing have been performed on the binary hand silhouette to obtain the perfect shape of the gesture.

2.2 COLOR SPACE MODELS

Colour is the most important information in segmentation, gesture and object recognition and in many image processing applications. Skin colour segmentation is one of the mostly employed segmentation process in gesture recognition. Skin region detection is the primary step in gesture as well as face detection. Most of the skin colour detection algorithms are based on colour space models [1]. Colour space models are used to represent images in three or four primary colour spaces. There are mainly four colour space models which are used to detect skin colour regions in a gesture or a face. They are RGB colour space model, YCbCr colour space model, YIQ colour space model and HSV colour space model. In pattern recognition and image processing applications, choosing a perfect colour space model is of paramount importance because some of the original colours in the image might not be suitable for the particular application. Details study of several colour space models and their comparisons for skin colour detection is given in [1]. Here we will discuss four primary colour space models.

2.2.1 RGB COLOR SPACE MODEL

RGB colour space [6] consists of three additive primary colours: red, green and blue. RGB colour space can produce any colour which can be made by the combination of these three primary colours. RGB colour space works similarly as human visual system. For that reason, it is widely used colour space model in Computer Vision.

The RGB colour model is represented by a three dimensional cube with red green blue at the corners of each axis as shown in Fig. 1. In RGB colour model red, green and blue components have a high level of correlation. For that reason, RGB colour space is not a very good choice for skin colour segmentation.

To reduce the effects of illumination and light intensity on different images, the R,G and B values are normalized by a simple normalization technique as given below

$$r = \frac{R}{R+G+B} \tag{2.1}$$

$$g = \frac{G}{R+G+B} \tag{2.2}$$

$$b = \frac{B}{R+G+B} \tag{2.3}$$

Here, the sum of the normalized values of the R, G and B is unity.

$$(r+g+b)=1 \tag{2.4}$$

The normalized RGB colour space model is more popularly used in skin color detection than RGB colour space model, because of its illumination invariance.

2.2.2 HSI COLOR SPACE MODEL

HSI colour space [7] is a combination of HSL and HSV colour spaces. HSL and HSV colour models both are cylindrical coordinate representation of points of RGB colour space as given in Fig.2. Both HSL and HSV colour spaces are perceptually more reliable because of its cylindrical representation rather than Cartesian representation.

In HSI the term H corresponds to hue value of the colour, S corresponds to saturation value of the colour and I corresponds to intensity value of the colour. In HSL and HSV L and V stand for luminance and brightness respectively. Although HSL, HSV and HSI models are widely used in image processing applications, these models are not perceptually uniform.

HSL and HSV both have cylindrical geometry. Where hue is the angular dimension and it starts at red component at 0 degree, passes through green component at 120 degree and blue component at 240 degree, as given in Fig.2.

The main advantage of HSI colour space is in this colour space, we don't need to know the values of green and blue components. To change a deep red value to pink, by changing the saturation value only we can adjust. In machine vision and image processing HSI colour space has a huge application. Some researchers have proposed skin colour segmentation based on HSI colour space in [3].

2.2.3 YCbCr COLOR SPACE MODEL

YCbCr colour space [1] is mainly used in European television studios. It can represent colours with statistically independent components. For that reason, it results in a uniform clustering of colours. RGB colour space has a redundancy in its colour components, and it doesn't show uniform clustering of colour components. In YCbCr, Y represents the luminance component of colour, Cb and Cr represent chrominance between two colours blue-yellow and red-yellow respectively.

YCbCr has a uniform separation between chrominance and luminance. For that reason, it is the most popular colour space model for skin colour detection.

2.2.4 YIQ COLOR SPACE MODEL

YIQ colour space is derived from YCbCr colour space. Here Y corresponds to luminance value, I and Q correspond to chrominance value. For that reason, it is also a widely used segmentation process. Y represents the luminance value and I and Q represent chrominance value. It represents the change from orange to cyan, where Q represents the change from purple to yellow- green. This colour space separates luminance and hue information. For that reason, it is also widely used in skin colour detection.

Researchers have proposed hand gesture segmentation using YIQ colour space model in [4].

2.3 YCbCr COLOR SPACE BASED SKIN COLOR SEGMENTATION

. Skin colour segmentation using YCbCr colour space model is very popular because it can represent colours with statistically independent components. For that reason, it results in a uniform clustering of colors. Other color spaces like RGB color space have a redundancy in its color components, and it doesn't show uniform clustering of color components. We have used YCbCr color space based skin color detection.

In this step, skin color region of the hand gesture is segmented using YCbCr skin color segmentation. We have made our database so that the hand region has the maximum area compared to the other objects.

We captured our images in RGB color space. To make the algorithm illumination normalized, the R, G and B values are normalized by dividing each value by the sum of R, G, and B components.

Then the normalized RGB images are converted into YCbCr color [1] images by the following formula [1].

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 65.738 & 129.057 & 25.064 \\ -37.945 & -74.494 & 112.439 \\ 112.439 & -94.154 & -18.285 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.5)$$

The threshold value of Cb, Cr and Y is proposed for skin color segmentation as 85<Cb<128, 129<Cr<185 and Th<Y<255. Where Th is the 1/3th of the mean value of Y component.

Skin color segmentation using YCbCr color space is described in Fig.3.

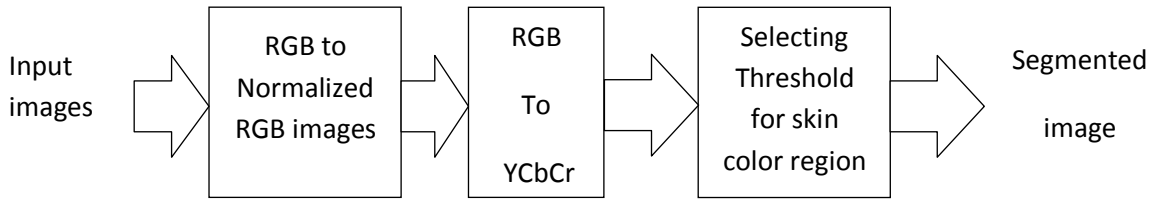


Figure 2. 1 YCbCr color space based skin color segmentation

2.4 THEORY OF K-MEANS CLUSTERING

Clustering is a supervised or unsupervised learning method by which any object or data can be partitioned into two or more groups. These groups are called clusters. Many researchers have employed clustering techniques in segmentation. Segmentation is the process to segment an image into foreground and background. So, by properly choosing clustering criterion segmentation can be performed. Labeling is the most difficult problem in clustering. If we do not use labeling in clustering, it will be called unsupervised clustering otherwise it is called supervised clustering.

K-means clustering is the most popular clustering method among the all clustering methods as it is very simple to implement. It is not an inclusive clustering algorithm like Fuzzy C-means clustering. In K-means clustering any point can belong to only one cluster. It is an iterative technique, by which any image or data can be partitioned into two or more regions.

The minimizing objective function is given by [8]

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} (\mathrm{x}_i - \mu_i)^2 \tag{2.6}$$

Here number of cluster is k, μ_i is the centroid of the ith point.

The basic algorithm is:

i) Assign some initial centroids to i number of points.

ii) Depending on the equation (2.6), assign each object to a group which has the nearest centroid.

iii) After assigning all points, relocate the K centroids. The new centroids for each cluster are calculated by the following formula (2.8).

iv) Repeat step ii) and step iii) until the algorithm converges.

$$c^{(i)} = \arg\min \| x^{(i)} - \mu_j \|^2 \tag{2.7}$$

$$\mu_i = \frac{\sum_{i=1}^{m} \{c_i = j\} x^{(i)}}{\sum_{i=1}^{m} \{c_i = j\}} \tag{2.8}$$

It produces simple foreground background separation of an image. For that reason this algorithm is very popular in uniform background subtraction.

Although it can be proved that this algorithm will always converge, it has some disadvantages:

1 Initial allocation of centroids in K-means algorithm tends to wrong partition of clusters.

2 Many times a convergence problem occurs because of empty cluster generation.

3 It may include a small cluster within a large cluster.

For that reason, many times segmentation result is not satisfactory.

To overcome this problems, we have proposed modified k-means clustering [9] as described in next section.

2.5 MODIFIED K-MEANS CLUSTERING AND MAHALANOBISH DISTANCE

Modified K-means algorithm: Initial allocation of centroids in K-means algorithm tends to wrong partition of clusters. Sometimes a null set can be treated as a cluster. This null set is called empty cluster. The centre updatation process in the k-means algorithm is given by the following formula [8]

$$z_k^{(new)} = \frac{\sum_{x_j \in c_k} (x_j)}{n_k} \tag{2.9}$$

Where n_k is the number of elements in cluster c_k. If the new centres $z_k^{(new)}$ do not match exactly with the old centres $z_k^{(old)}$, the k-means algorithm enters into new iteration assuming $z_k^{(new)}$ as $z_k^{(old)}$. Because of this iteration process sometimes empty clusters have been generated in k-means clustering.

To avoid the empty cluster generation researchers have proposed a new modified k-means clustering in [9]. This algorithm is same as original k-means algorithm except the new centroid allocation step. In this algorithm centroids for new clusters are generated by the following formula [9]

$$z_k^{(new)} = \frac{\sum_{x_j \in c_k} (x_j) + z_k^{(old)}}{n_k + 1} \tag{2.10}$$

In this scheme as we consider new clusters as member of previous cluster, empty cluster generation is totally avoided.

Mahalanobish distance: The Mahalanobish distance [11] is a statistical measure, used to analyse the similarity between an unknown and a known data set. Mahalanobish distance was described by famous mathematician P. C. Mahalanobish. Mahalanobish distance includes mean and covariance matrix of the dataset.

Let we have two datasets X and Y. Where, $X = (x_1, x_2, x_3 ..... x_n)^T$ is a known dataset and $Y = (y_1, y_2, y_3 ..... y_n)^T$ is an observation or unknown dataset. Mahalanobish distance measures similarity or dissimilarity between these two datasets by the following equation

$$d_M(x, y) = \sqrt{(x_i - y_j)^T S^{-1} (x_i - y_j)} \tag{2.11}$$

Here, $d_M(x, y)$ is the Mahalanobish distance between these two datasets. S is the covariance matrix of the known dataset X.

Modified K-means algorithm performs well for uniform background. By applying modified K-means algorithm on database 1, we have extracted skin color regions. We have proposed a semi-supervised learning algorithm based on Mahalanobish distance, to find the hand regions or foreground regions in the complex background database. We found the Mohalanobish distance [11] of images with the extracted skin color regions of database 1. Our proposed algorithm has been described in next chapter.

2.6 PROPOSED ALGORITHM FOR HAND GESTURE SEGMENTATION

We have employed a semi-supervised learning algorithm based on modified K-means clustering [9] and mahalanobish distance [11] to extract the skin color region from the captured hand gesture images. The proposed algorithm is described below:

Step 1: Convert the RGB images of database 1 (uniform background) to YCbCr color space images.

Step 2: Reshape the images in Y, Cb and Cr components.

Step 3: Perform modified K-means clustering with cluster size 2. Let [a1,b1]=m_kmeans(image1,2).

Step 4: Assuming hand region has the minimum area in all the images, find out hand region from the foreground. Hand=b1 (minimum_area).

Step 5: Reshape the hand region in in Y, Cb and Cr components and Perform modified K-means clustering with cluster size 2. ROI=reshape (Hand,[],3). [a2,b2]= m_kmeans (ROI,2).

Step 6: Convert the RGB color images of database 2(complex database) to normalized RGB images, to reduce the illumination effect.

Step 7: Convert normalized RGB images to YCbCr images.

Step 8: Find Mahalanobish distance (d) between reshaped data obtained from second database and centroids of clusters of hand regions obtained from first database.

Step 9: Perform modified K-means clustering on d, with cluster size 2. Let [a3,b3]=m_kmeans(d,2)

Step 10: Here b3 consists of only two values, 1 and 2.1 value corresponds to 1^{st} cluster (foreground) and 2 value correspond to 2^{nd} cluster (background). Replace all the 2 values by 0. Thus we get a binary silhouette of hand gesture.

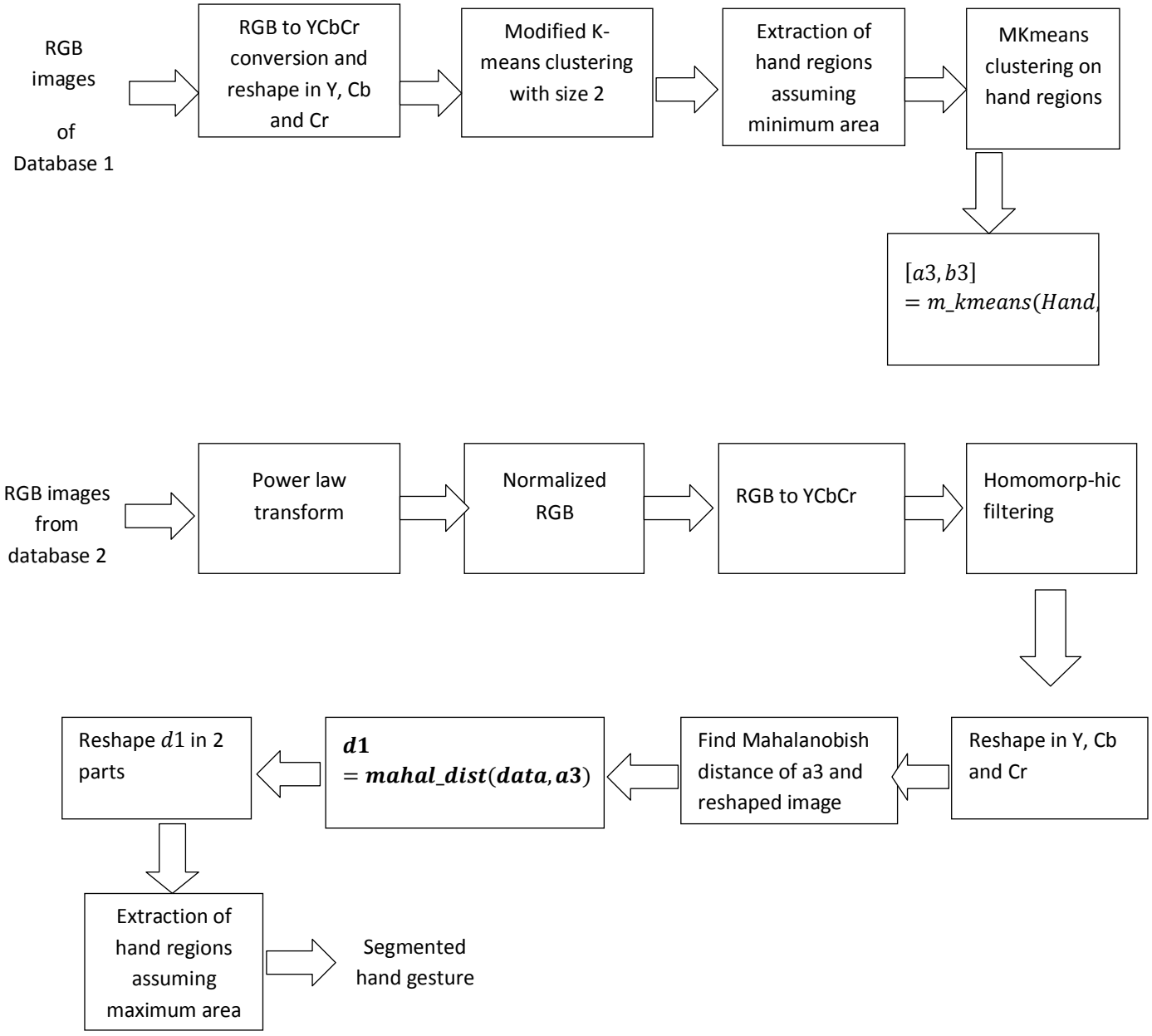


Figure 2. 2 Block diagram of proposed Segmentation process

2.7 ILLUMINATION AND ROTATION NORMALIZATION

We captured our images in various angles to make the system rotation variant. Because of this rotation angle variation, gestures of the same classes might be misclassified as gestures of other classes. To make our gesture recognition algorithm rotation invariant we have employed a rotation invariant algorithm proposed in [4]. Because of the variations in light intensity our images are illumination variant. We have proposed some methods to make our gesture recognition algorithm illumination invariant.

2.7.1 ROTATION INVARIANT ALGORITHM

Researchers have proposed an algorithm to make all the gestures rotation invariant in [4]. In this method, direction of principal axes and the rotation angle between gesture and principal axes is found. Then the segmented gesture is rotated to coincide the principal axis and vertical axes.

Fig.3 shows how effectively this algorithm has made our segmented gestures rotation invariant.

2.7.2 ILLUMINATION INVARIANT ALGORITHM

To make our gesture recognition algorithm illumination invariant, we have implemented three steps: a) Power law transform on images, b) converting RGB images to normalized RGB space and c) homomorphic filtering.

A) Power law transform: Power law transformation [10] is expressed by the following equation

$$s = c\, r^{\gamma} \tag{2.12}$$

Where, c and γ are positive constants. Here γ is called Gama constant, and it is used to control the intensity values of an image. When $o < \gamma < 1$ all the dark input values are transformed into a wider value. Thus with fractional values of γ illumination level of dark images increases. When $\gamma > 1$ opposite effect occurs.

In this work we have empirically selected γ as 0.5, to increase overall illumination level.

B) RGB to normalized RGB: We have captured our images in RGB color space. Some researchers have proposed an illumination normalization technique by converting RGB images to normalized RGB images [4].

We converted our power law transformed RGB colour images to normalized colour images.

C) Homomorphic filtering: Any image f (x, y) can be expressed as a product of illumination and reflection components as given below [10]

$$im(x,y) = il(x,y)ref(x,y) \tag{2.13}$$

Here il(x, y) is the illumination and ref (x, y) is the reflectance. In homomorphic filtering intensity range is compressed and contrast is enhanced by a frequency domain process as described in the following figure.

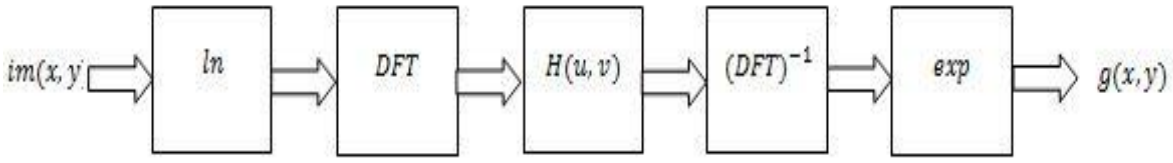


Figure 2. 3 block Diagram of Homographic Filtering

Equation 1 cannot be used directly on the illuminance and reflectance components in frequency domain because

$$\mathrm{F}(\mathrm{im}(\mathrm{x},\mathrm{y})) \neq \mathrm{F}(il(x,y))\,\mathrm{F}(\mathrm{ref}(\mathrm{x},\mathrm{y}))$$

For that reason, we have used logarithm on the product of illumination and reflectance values. Then DFT operation is performed on the log transformed image, as given in the following formula

$$\mathrm{zz}(\mathrm{x},\mathrm{y}) = \ln(\mathrm{im}(\mathrm{x},\mathrm{y})) = \ln(il(x,y)) + \ln(ref(x,y))$$

$$\mathrm{F}(\mathrm{zz}(\mathrm{x},\mathrm{y})) = \mathrm{F}(\ln(\mathrm{im}(\mathrm{x},\mathrm{y}))) + \mathrm{F}(\ln(ref(x,y))) \tag{2.14}$$

$$\mathrm{ZZ}(\mathrm{u},\mathrm{v}) = \mathrm{FF}_i(u,v) + FF_r(u,v) \tag{2.15}$$

Where, $FF_i(u,v)$ is the Fourier transform on $\ln(il(x,y))$ and $FF_r(u,v)$ is the Fourier transform on $\ln(ref(x,y))$.

We use a filter of transfer function $HH(u,v)$ to filter $ZZ(u,v)$.

Now, $SS(u,v) = HH(u,v)ZZ(u,v) = HH(u,v)FF_i(u,v) + HH(u,v)FF_r(u,v)$Now, using inverse DFT, we can get

$$ss(x,y) = F^{-1}(SS(u,v) = F^{-1}\{HH(u,v)FF_i(u,v)\} + F^{-1}\{HH(u,v)FF_i(u,v)\} \quad (2.16)$$

Let, $i'(x,y) = F^{-1}\{HH(u,v)FF_i(u,v)\}$ and $r'(x,y) = F^{-1}\{HH(u,v)FF_r(u,v)\}$ (2.17)

So, we can say that $ss(x,y) =, i'(x,y) + r'(x,y)$ (2.18)

Here, $i'(x,y)$ contains the logarithmic part of illuminance component and $r'(x,y)$ contains the logarithmic part of reflectance component.

By taking exponential on both $i'(x,y)$ and $r'(x,y)$ we will get the original illuminance and reflectance component.

$io(x,y) = e^{i'(x,y)}$ and $ro(x,y) = e^{r'(x,y)}$ (2.19)

The main idea behind homomorphic filtering is to separate illumination and reflectance components which is described by the above mathematical expressions. There is an interesting phenomenon in image, illumination component of an image vary very slowly whereas, reflectance component vary very abruptly in spatial domain. So, we have to control both these components to manage the illumination variations in an image. It is done by properly choosing the transfer function of the homomorphic filter. Transfer function of the homomorphic filter is given by

$$HH(u,v) = (\gamma_H - \gamma_L)\left[1 - e^{-c\left[\frac{D^2(u,v)}{{D_0}^2}\right]}\right] + \gamma_L \quad (2.20)$$

If the parameters $\gamma_H$and γ_L chosen so that, $\gamma_L < 1$ and $\gamma_H > 1$, it will increase the high frequency part(reflectance component) and decrease low frequency part (illuminance component).

In our work, we have selected $\gamma_L$as 0.8, γ_H as 1.2 and D_0 as 20. The constant c is chosen 1.

2.8 MORPHOLOGICAL OPERATIONS

From fig. 5 it is clear that our proposed two segmentation algorithms are not sufficient to find out the region of interest, binary hand silhouette. In the segmented image of fig. 5 there are some background and object noises, which are undesirable in gesture recognition. In this work, we have employed four fundamental morphological operations: a) erosion b) dilation c) opening and d) closing to obtain the proper shape of the binary silhouette of hand gestures [10].

Morphology is the study of the form and structure of plants, animals, bacteria, virus and many other living organisms. In image processing similar analogy is used to find out the original object shape, it is called mathematical morphology. Mathematical morphological operations can be employed in various ways. Even it can be used in n-dimensional fields. Here we will discuss only those morphological operations which are used to find out shape of an object from a binary image.

Mathematical morphological operations are solely derived from set theory. In morphological operations various objects within an image are represented by different sets. Two basic concepts of set theory namely set reflection and set translation are the basics of morphological operations.

The reflection of a set B is defined as B_r [10]

Where B is the set of pixels which represents an object in the image and B_r is the set of points whose (x, y) coordinates are replaced by (-x,-y).

The translation of set A by an arbitrary point y= (y1, y2) is given by the following equation

where B_y is the set of points in A whose (x, y) coordinates are replaced by (x+y1,y+y2).

These two are the fundamental set theory operations which are used in morphological operations. Now we will discuss erosion, dilation, opening and closing operations in brief.

EROSION: The erosion of two sets A and B , $A \odot B$ is the intersection of set A and set $(B)_z$. Where, $(B)_z$ is the translation of set B by a point z. Here set B is the structuring element and A is the original image. By selecting a proper structuring element, we can perform erosion operation on an image.

DILATION: The dilation of two sets A and B, $A \oplus B$ is the intersection of set A and $\hat{B}$, where, $\hat{B}$ is the reflection of set B. By selecting a proper structuring element, we can perform dilation operation on an image.

Dilation is used to expand the components of an image and erosion is used to shrink the components of an image by choosing proper structuring elements.

OPENING: The opening of a set A by another set B is given by the following equation [10]

$$A \circ B = (A \odot B) \oplus B \tag{2.21}$$

So, the opening of A by B, $A \circ B$ is the erosion of A by B, followed by a dilation by B. Here set B is the structuring element and A is the original image. By selecting a proper structuring element, we can perform opening operation on an image.

Opening operation normally smoothens image boundary and separates any kind of object noises from the original image.

CLOSING: The closing of a set A by another set B is given by the following equation [10]

$$A \bullet B = (A \oplus B) \odot B \tag{2.22}$$

So, the closing of A by B, $A \bullet B$ is just opposite of opening operation. Closing is the dilation of A by B, followed by erosion by B. Here set B is the structuring element and A is the original image. By selecting a proper structuring element, we can perform closing operation on an image.

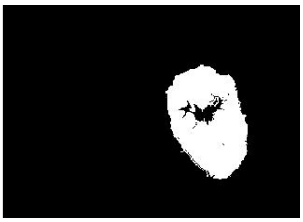
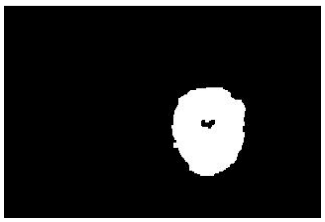
Opening operation also smoothens image boundary and separates any kind of object noises from the original image like opening operation but it generally removes the narrow breaks and long thin gulfs and fills all the gaps in an image contour.

In our segmentation process we have used a sequence of dilation, erosion, opening and closing operations to get binary hand silhouette.

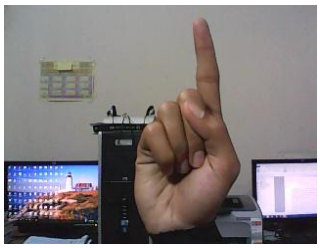
2.9 RESULTS AND DISCUSSIONS

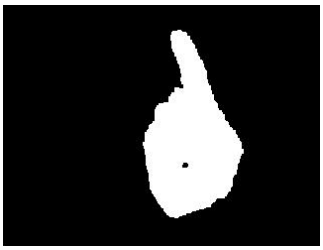
We have done segmentation of static hand gestures using YCbCr skin color based segmentation and modified K-means clustering based semi-supervised learning. Here in this part, we will discuss segmentation result of both these methods.

YCbCr skin color detection: In this method skin color regions are detected from the original image by a pre-defined threshold of Y, Cb and Cr. Assuming hand region is of maximum area, we have extracted the hand region. Rotation and illumination normalization and some morphological operations are done on the hand region, to extract the region of interest. The overall results of this segmentation process have been described in following figures.

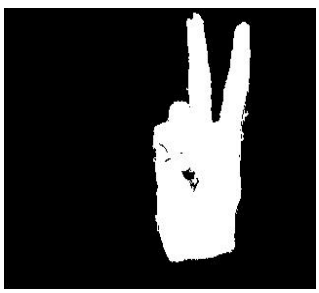
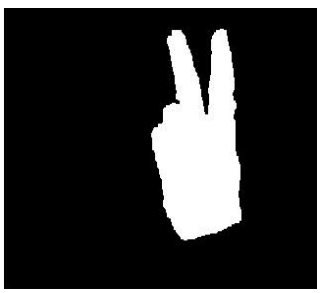




Original Image of digit "0" Segmented Image Rotation invariant and morphological operation

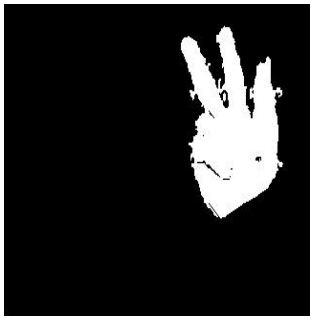
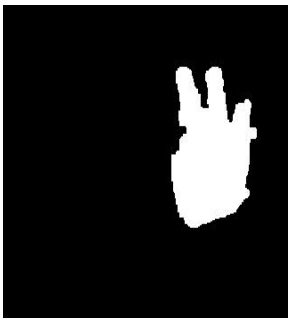




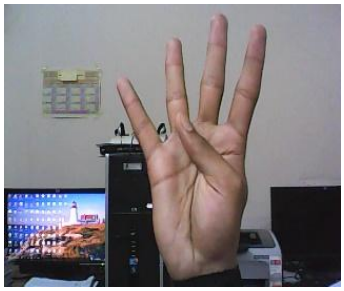
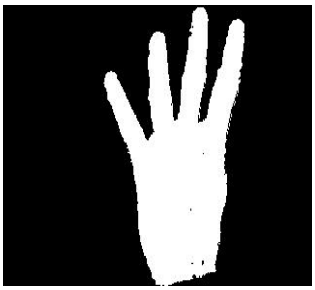
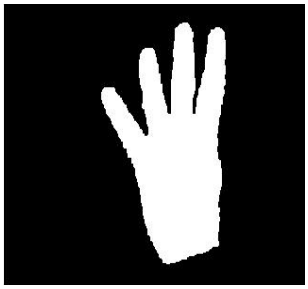
Original Image of digit "1" Segmented Image Rotation invariant and morphological operation

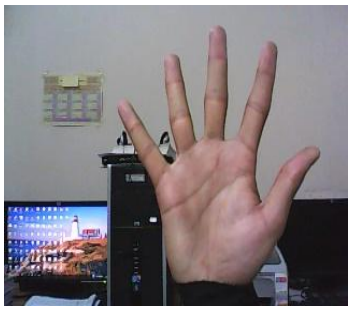
Original Image of digit "2" Segmented Image Rotation invariant and morphological operation

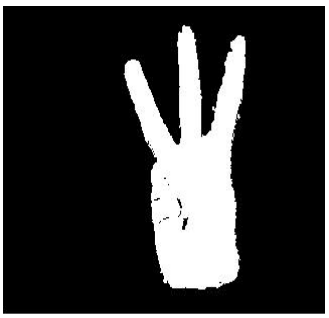
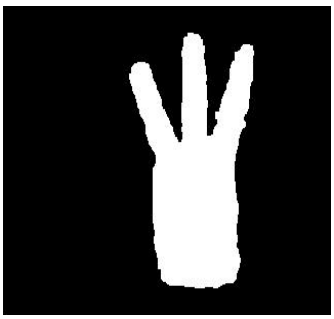
Original Image of digit “3” Segmented Image Rotation invariant and morphological operation

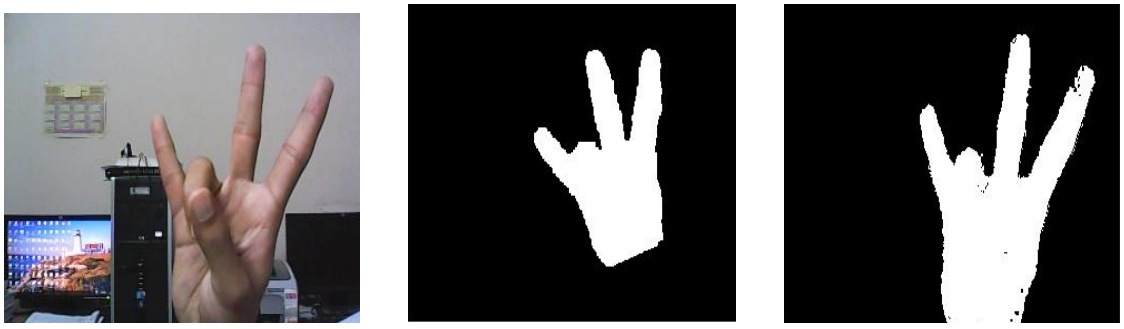
Original Image of digit “4” Segmented Image Rotation invariant and morphological operation



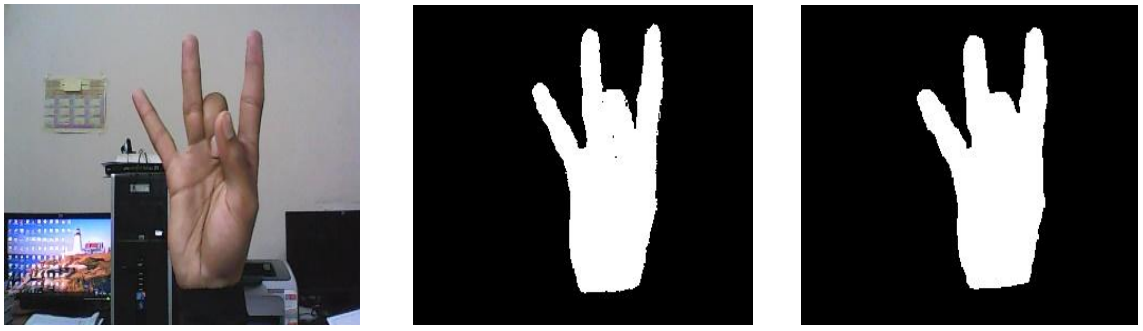
Original Image of digit “5” Segmented Image Rotation invariant and morphological operation

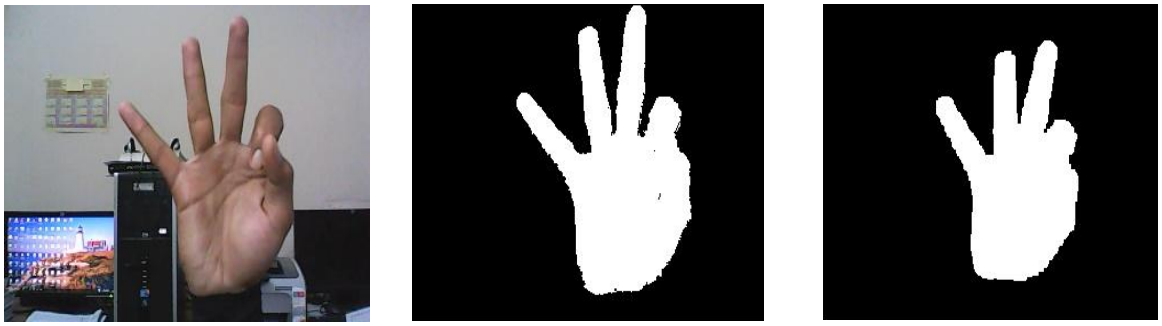
Original Image of digit “6” Segmented Image Rotation invariant and morphological operation



Original Image of digit “7” Segmented Image Rotation invariant and morphological operation



Original Image of digit “8” Segmented Image Rotation invariant and morphological operation

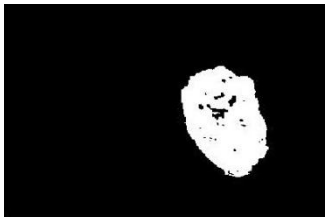
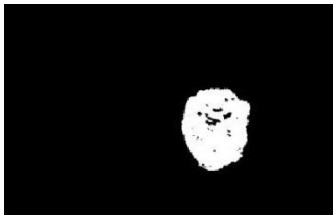


Original Image of digit “9” Segmented Image Rotation invariant and morphological operation

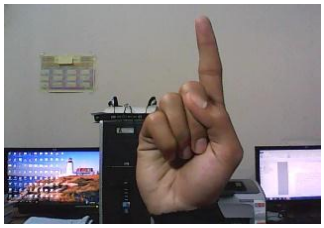
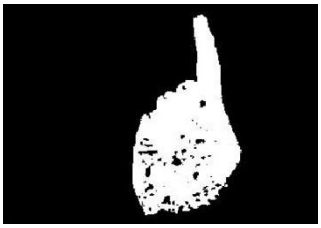
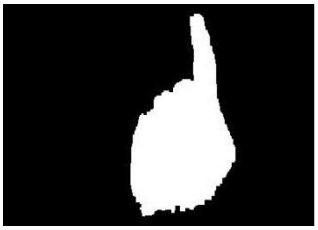
Figure 2. 4 Overall results of this Segmentation Process

Segmentation based on semi-supervised learning : In this method we have used our first database (uniform background) to find out skin color regions by an unsupervised learning method called modified K-means clustering. Then, we have calculatd mohalanobish distance to extract the hand region from our second database(complex background). This method has shown illumination invariance and has given better segmentation result than previously used

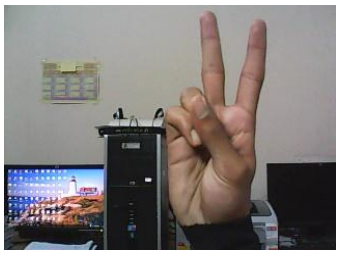
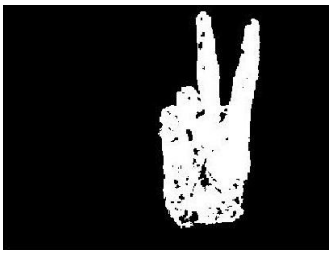
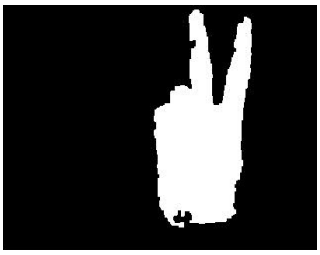
YCbCr based skin color segmentation method. Segmentation results are shown in the following figures.

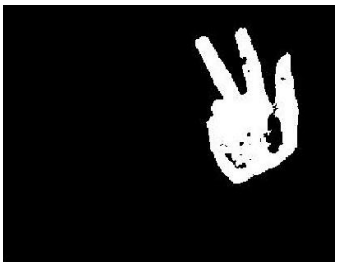
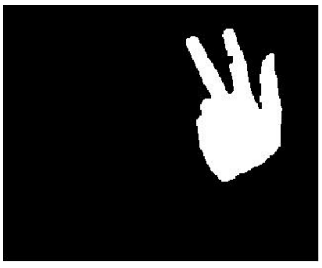
Original Image of digit “0” Segmented Image Rotation invariant and morphological operation

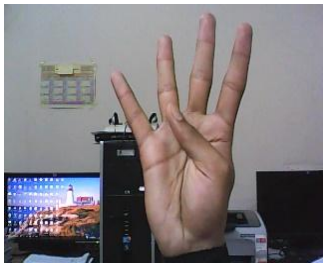
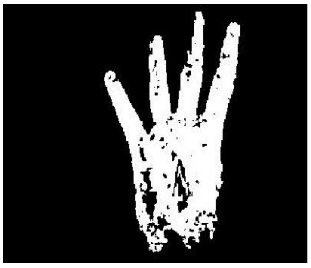
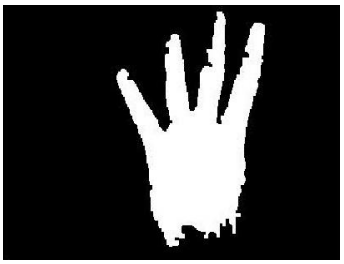
Original Image of digit “1” Segmented Image Rotation invariant and morphological operation

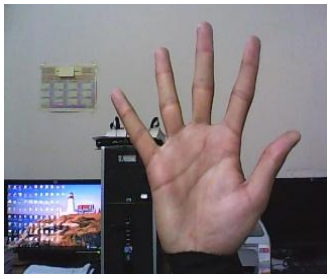
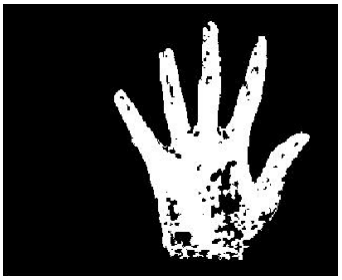
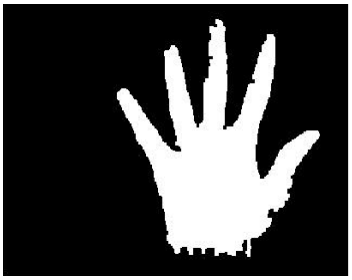
Original Image of digit “2” Segmented Image Rotation invariant and morphological operation

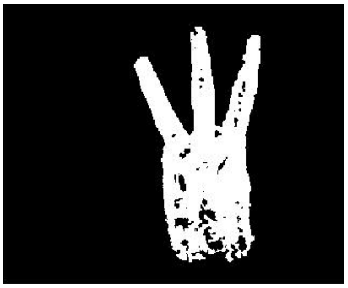
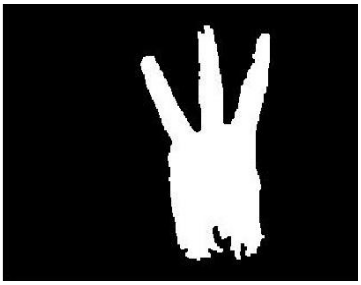
Original Image of digit “3” Segmented Image Rotation invariant and morphological operation

Original Image of digit “4” Segmented Image Rotation invariant and morphological operation

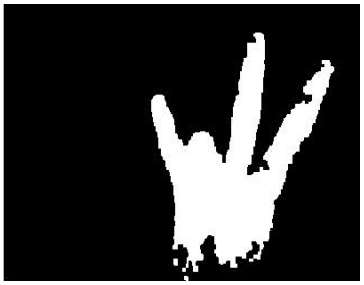




Original Image of digit “5” Segmented Image Rotation invariant and morphological operation

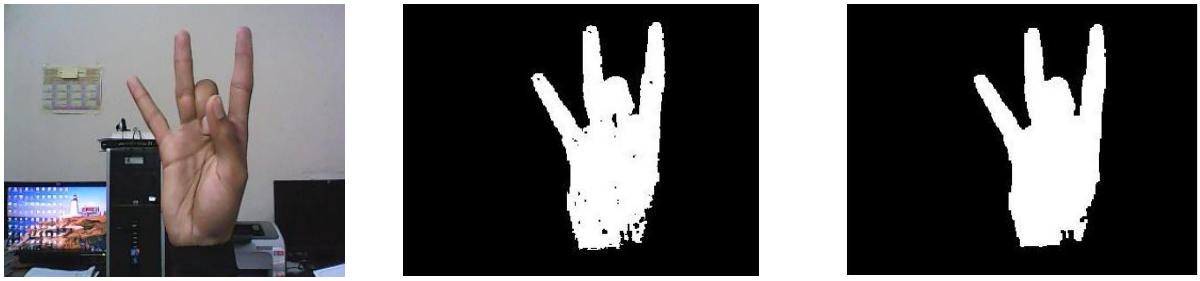




Original Image of digit “6” Segmented Image Rotation invariant and morphological operation

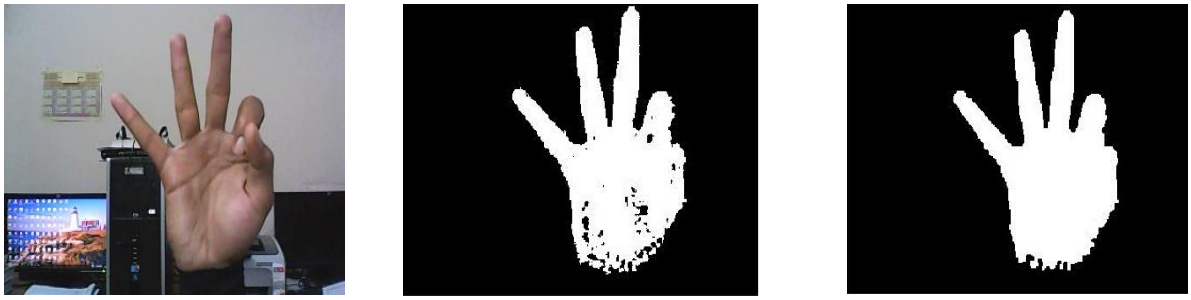




Original Image of digit “7” Segmented Image Rotation invariant and morphological operation



Original Image of digit “8” Segmented Image Rotation invariant and morphological operation



Original Image of digit “6” Segmented Image Rotation invariant and morphological operation

Figure 2. 5 Results of Segmentation based on semi-supervised learning

2.10 CONCLUSIONS

Segmentation is the primary and basic step for gesture recognition. Skin colour region detection in complex background and varying illumination condition is a difficult task for researchers. Most of the researchers have employed colour space based skin colour segmentation for Static Hand Gesture Recognition. Colour-space base skin colour segmentation methods are not robust for skin colour detection, because in varying illumination condition and in complex background, threshold values for the colour space models also vary. We have proposed a skin colour detection process using semi-supervised learning based on K-means clustering and Mahalanobish distance, which has shown robustness in varying illumination condition and complex background.

Our proposed illumination normalization technique has shown effectiveness in varying illumination condition. We have employed a rotation normalization technique, proposed by some researchers. This rotation normalization technique has also shown robustness with the changes of orientation of hand.

REFERENCES

1. H. B Ghazali., J. Ma and R. Xiao "An Innovative Face Detection based on Skin Color Segmentation*" International Journal of Computer Applications* (0975 – 8887), Vol. 34, No.2, pp. O975- 8887, Nov. 2010.
2. K. Muthukannan and M. M. Moses **"**Colour image segmentation using k-means clustering and Optimal Fuzzy C-Means clustering*" International Conference on Communication and Computational Intelligence (INCOCCI),* 2010, Erode, India, Dec. 2010.
3. S. P. Priyal and P. K. Bora "A Robust Static Hand Gesture Recognition System Using Geometry based Normalization and Krawtchouk Moments*" Pattern Recognition*, vol. 46, no.8, pp. 2202-2219, 2013.
4. D. K. Ghosh and S. Ari "A Static Hand Gesture Recognition Algorithm Using K-Mean Based Radial Basis Function Neural Network" *In: 8th International Conference on Information, Communications and Signal Processing (ICICS),* pp. 1-5, Singapore , 2011.
5. T.William, T. Freeman and M. Roth "Orientation histogram for hand gesture recognition" *in proceedings of the 1st International workshop on Automatic face and gesture recognition, pp. 296- 301,* 1995.
6. V. E. C. Ghaleh. and A. Behrad "Lip contour extraction using RGB color space and fuzzy c-means clustering" *9th International Conference on Cybernetic Intelligent Systems (CIS),* Sep.2010, Reading, Berks, UK
7. E. Welch, R. Moorhead, and J. K. Owens, "Image processing using the HSI color space", *IEEE Proceedings of Southeastcon* '91, pp. 722 – 725, vol.2, Williamsburg, VA, April.
8. 2011. Haykin, S.: Neural networks. *Prentice-Hall (1999, 2nd edn.).*
9. M. K. Pakhira, "A Modified k-means Algorithm to Avoid Empty Clusters", *International Journal of Recent Trends in Engineering*, vol. 1, no. 1, May 2009.
10. R. C. Gonzalez. R. C. Digital Image Processing, *Pearson Education India.(2009).*
11. Y. Chen, Q. Wu, X. He, W.Jia and T. Hintz "A Modified Mahalanobis Distance for Human Detection in Out-door Environments*" First IEEE International Conference on Ubi-Media Computing*, Lanzhou, China, Jul. 2008.

CHAPTER 3

FEATURE EXTRACTION AND CLASSIFICATION

3.1 INTRODUCTION

Feature of an image can be thought as mathematical representation of binary silhouettes or image contours. Features are used to distinguish images of different classes. There are mainly two types of features of images a) shape-based features [1, 2, 3] b) contour based features [4, 5]. Shape-based features are calculated based on the intensity or pixel values of segmented hand gesture image. On the other hand, contour-based features are calculated on the boundary pixels or on the boundary profiles. Researchers have employed both these two feature extraction techniques in [1, 2, 3, 4, 5].

Contour based feature extraction techniques [4, 5] misclassify gestures of almost similar shapes, because gestures of similar shapes have almost similar contours and boundaries. In case of shape-based feature extraction techniques [1, 2, 3], this problem doesn't occur, because it is calculated on the whole image shape. For that reason shape-based feature extraction techniques are more preferable than contour-based feature extraction techniques.

In this work, we have used shape-based features to extract desired information from the segmented hand regions. Non-orthogonal moments like geometric moments [1] and orthogonal moments like Tchebichef [3] and Krawtchouk moments [1] are used here as features. All these moments are discrete. We have not used continuous moment like Zernike moment [1], because continuous moments use to produce discretization error, which is very much undesirable for pattern recognition.

In user-independent gesture recognition, many times none of the popularly used features shows satisfactory classification performance. To improve the performance of classification, two feature fusion strategies [6] are proposed in this work; serial feature fusion and parallel feature fusion. Feature fusion is the process by which we can combine two features based on some proper algorithm or statistical criterion. Here we have proposed two very simple techniques, which are effective for both same size and different size features.

Classification is the final stage in pattern recognition. We have designed an Artificial Neural Network (ANN) [7]. It is trained and tested using the feature sets described in chapter 3. The dataset consists of 1500 color images of 10 gestures, 15 sample each class of 10 users. The dataset is equally divided into training and testing dataset of 750 images for 5 different users both to make the system user-independent.

The orders of Geometric moment, Krawtchouk moment and Tchebichef moments are empirically chosen as 49(n=7,m=7), 64(n=8,m=8) and 64(n=8,m=8) respectively. The parameters p1 and p2 of Krawtchouk moments are set to 0.5 each. In ANN classification we have empirically chosen learning rate parameter, momentum constant, number of epochs 0.5, 0.9 and 10000 respectively. We have used tansigmoid activation function. We have used single hidden layer and number of nodes in hidden layer is empirically chosen 200. Number of nodes in input layer equals to feature size(49, 64 and 64 for geometric, krawtchouk and tchebichef moments respectively). Number of nodes in output layer equals to number of classes. A two-fold operation is done to evaluate the generalized performance of the system in user-independent condition.

Design: We have designed a feed-forward multi-layer perceptron (MLP) neural network with a single hidden layer for classification [7]. The number of neurons in the hidden layer is empirically set to 200. The number of input nodes is equal to the number of features. So the number of input nodes is 64, 64 and 49 respectively for Krawtchouk, Tchebichef and Geometric moments respectively. Number of output nodes is equal to number of classes. Here we have used 10 classes, so number of output nodes are 10.

TRAINING: The network was trained using 750 color images of 10 gestures, 15 samples of each class of 5 users.

TESTING: A total of 750 hand gesture images of 10 classes, 15 samples of each class of 5 users were tested. Here we have used totally different 5 users to make the system user-independent.

TWO-FOLD OPERATION: A two-fold operation is performed to evaluate the generalized performance of the system in user-independent condition. In two fold operation, testing and training dataset are altered, and the average values of all the performance parameters are taken.

In Section 3.2 the details of the proposed feature extraction techniques are given. Feature fusion techniques are discussed in Section 3.3, Section 3.4 has shown the results of feature extraction and feature fusion techniques. Section 3.5 concludes the paper.

3.2 THEORY OF MOMENT FEATURES

In Pattern Recognition, moments [1] are widely used shape-based features for images. Image moment is calculated as a weighted average of image pixels. It is a global as well as local representation of image features. Moments are popularly chosen as a shape-based feature, because moment features are rotation and scale invariant. Moment features are normally of two types, continuous and discrete. In Pattern Recognition, normally continuous moments are not used because in continuous moments discretization error occurs. Though, some continuous moments like Zernike moments [1] are several times used as a feature in image processing. There are also two types of moments: orthogonal and non-orthogonal moments. Orthogonal moments are expressed as a product of two mutually orthogonal functions. On the other hand, non-orthogonal moments are calculated on the image only. They don't have any orthogonal functions. Some orthogonal moments like Krawtchouk [1] moments have normalization factor, but most of the moments don't have any normalization factor.

Here we have used orthogonal and non-orthogonal Moment features [1]. For an image $f(x,y)$ of size N×N with $(x,y) \in \{0,1,2.....N-1\} \times \{0,1,2.....M-1\}$, the moments are given by the following equations [1].

Geometric moment: The geometric moment is described by the following equation [1]

$$GM_{rs} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} x^r y^s f(x,y) \tag{3.1}$$

Here order of the moment is (r+s). n=0……..N-1, and m=0……M-1.

The corresponding central moment of order (r+s) is defines as [7]

$$\mu_{rs} = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} \left(x-\bar{x}\right)^r \left(y-\bar{y}\right)^s f(x,y) \tag{3.2}$$

Where, r= 0,1,2,……M-1 and for s= 0,1,2,…….N-1

$$\bar{x} = \frac{GM_{10}}{GM_{00}} \tag{3.3}$$

$$\bar{y} = \frac{GM_{01}}{GM_{00}} \tag{3.4}$$

The normalized central moment is defined as …

$$\eta_{rs} = \frac{\mu_{rs}}{\mu_{00}^{\ \lambda}} \tag{3.5}$$

Where,

λ =scaling normalization factor=(r+s)/2+1. (3.6)

Krawtchouk moment: The Krawtchouk moments are one type of discrete orthogonal moments derived from the Krawtchouk polynomials .The nth order Krawtchouk polynomial at a discrete point x with (0<p<1,q=1-p) is defined in terms of hyper geometric function as [1]

$$k_1(x, p, N) = F\left(-n, -x - N, \frac{1}{n}\right) \tag{3.7}$$

By definition,

$$F(\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{z}) = \sum_{v=0}^{n} a_v\, b_v\, z^v / (c_v\, v!) \tag{3.8}$$

Where $(\mathrm{a})_v$ is the pochhammer function and is given by

$$(a)_v = a(a-1).....(a+v-1) \tag{3.9}$$

From the Krawtchouk polynomials we can get a weight function which is used to normalize feature values.

$$W(x, p, N) = \frac{\mathrm{N!}}{\mathrm{y!}(N-y)!}\, p^x\, (1-p)^{N-x} \tag{3.10}$$

Krawtchouk polynomials are normalized by dividing it by weight function as given in the following equation.

,where,

$$\mathrm{k}^1{}_1(x, p, N) = k_1(x, p, N)\sqrt{\frac{W(x, p, N)}{\rho(x, p, N)}} \tag{3.11}$$

$$\rho(n,p,N)=\frac{(-1)^n\left(\frac{1-p}{p}\right)^n n!}{(-N)_n} \tag{3.12}$$

The constant p is called shift parameter. Normally it is set to 0.5 to make the image centralized. Krawtchouk moment of order (n+m) is given by

$$Q_{nm}=\sum_{x=0}^{N-1}\sum_{y=0}^{M-1} f(x,y)k^1{}_1(x,p,N)\ k^1{}_2(\mathrm{y,q,M}) \tag{3.13}$$

Tchebichef moment: Tchebichef moments are same as Krawtchouk moment, as they are also derived from orthogonal basis functions. The 1D Tchebichef polynomial at a discrete point x is defined as [3].

$$\mathrm{t}_n(x)=(1-N)_n\,3_{F_2}(-\mathrm{n},-\mathrm{x},1+\mathrm{n};1,1-\mathrm{N};1) \tag{3.14}$$

Where, $3_{F_2}()$ is a hyper geometric function

$$3_{F_2}(\mathrm{a}_1,\mathrm{a}_2,\mathrm{a}_3;\mathrm{b}_1,\mathrm{b}_2;\mathrm{z})=\frac{(\mathrm{a}_1)_v(\mathrm{a}_2)_v(\mathrm{a}_3)_v z^v}{(\mathrm{b}_1)_v(\mathrm{b}_2)_v v!} \tag{3.15}$$

Where $(\mathrm{a})_\mathrm{v}$ is the pochhammer function and it is given by

$$(a)_v=a(a-1).....(a+v-1) \tag{3.16}$$

The Tchebichef moment of order (n+m) is given by [5]

$$T_{nm}=\frac{1}{\sqrt{\rho(\mathrm{n},N)\rho(\mathrm{m},N)}}\sum_{x=0}^{N-1}\sum_{y=0}^{M-1}\mathrm{t}_n(x)\mathrm{t}_m(y)\,f(x,y) \tag{3.17}$$

Where, $\rho(\mathrm{n},N)$ is normalization constant and is given by

$$\rho(\mathrm{n},N)=(2n)!\frac{(N+n)!}{(2n+1)!(N-n-1)!} \tag{3.18}$$

3.3 FEATURE FUSION

Fusion is the process of combining two or more objects or data to find the resultant which is more reliable than the originals. In Image processing and pattern recognition mainly three types of fusion techniques are used: Image fusion, Feature fusion and classification level fusion. We have used feature level fusion to increase classification performance.

In Pattern recognition research, feature fusion is the process of combining two or more features based on some pre-defined algorithm to extract relevant feature from one or more features. The resultant feature contains attributes of all the features. For that reason, feature fusion shows better classification performance than features. There are many techniques for feature fusion. Primarily cascading two features can make a feature level fusion as discussed in [10]. Some researchers have used Subspace learning to find the feature fusion as given in [12]. Many researchers have employed Canonical Correlation Analysis (CCA) for feature fusion [11]. None of the fusion strategies ensures increase in classification accuracy. We have used two feature level fusion strategies as described in [6].

To improve the accuracy rate in user independent situation, two feature fusion strategies [6]: parallel and serial feature fusion, have been proposed.

Let X and Y are two features consists of n and m numbers of feature vectors x_i and y_i. Here, $x_i \in X$ and $y_i \in Y$. Then the serial combined feature is defined as $F_s = [x\ y]$. The dimension of the resultant feature is (n+m). In case of parallel feature fusion, resultant feature is expressed by a super-vector $\mathrm{F}_p = [x_i + y_i]$. If the dimensions of X and Y re not same, then lower dimension vector is up-sampled to the dimension of upper-dimensional feature. Thus the size of the resultant parallel feature becomes m if $\mathrm{m} > \mathrm{n}$ or n if $n > m$.

For an example, let $X = [x_1\ x_2]$ and $Y = [\mathrm{y}_1\ y_2\ y_3]$. Then to find the parallel feature fusion of X and Y, we have to make X and Y of same length by up-sampling X to a length of 3. Now, $X = [x_1\ x_2\ x_3]$. Parallel combination is defined in a super plane as $\mathrm{F}_p = [x_i + y_i]$.

Numerical unbalance is the main problem in feature level fusion. Let a feature vector $\mathrm{X} = [0.5\ 0.9]$ and another is $\mathrm{Y} = [15\ 14]$. Then after parallel combination, attributes of Y will be more than X. It means almost all the significance of X will be lost. To overcome this

numerical unbalance condition, features have to be normalized first. Normalization of features is commonly done by dividing its maximum value as given below

$$\bar{X} = X / \max(X) \text{ and } \bar{Y} = Y / \max(Y). \quad (3.19)$$

Even after normalization, numerical imbalance may occur because of difference in feature sizes. To avoid this, lower dimensional one is multiplied by a normalization constant. We have empirically selected this normalization constant as given below

$$\nu = n^2 / m^2, \text{ assuming } n>m. \quad (3.20)$$

Then the serial and parallel feature vectors are given by

$$F_p = [x_i + \nu y_i] \text{ and } F_s = [x\ \nu y]. \quad (3.21)$$

Here, feature sizes of our primary features are 64, 64 and 49 for Krawtchouk, Tchebichef and Geometric moment features respectively. So the combination coefficient is unity for fusion of Krawtchouk and Tchebichef moments.

3.4 FEED-FORWARD MULTILAYER PERCEPTRON

Multilayer perceptron is a neural network with one or more hidden layers. It is a feed forward artificial neural network [11] that can be assumed as a connected graph between input, hidden and output nodes. Every nodes of multilayer perceptron is connected to each one except the input nodes. Each node in the multilayer perceptron is known as neuron. Back propagation learning algorithm is used in this types of artificial neural network.

Fig 4.1 shows the graphical representation of multilayer perceptron, which consists two

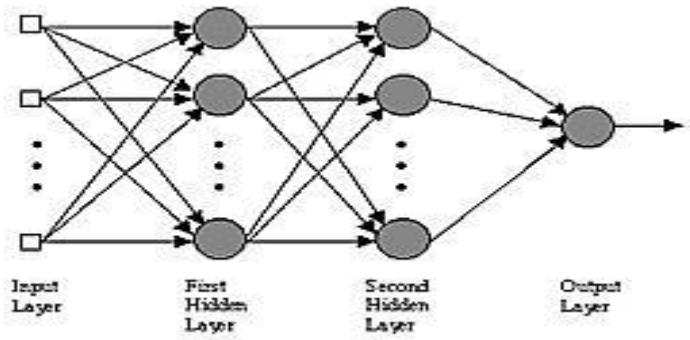


Figure 3. 1 Graphical representation of Multilayer Perceptron

hidden layers and one output layer. Here neurons are present in both the hidden and output layers. Hidden layer neurons take a very important role in the operation of multilayer perceptron. Hidden neurons are used to recover and extract features from the training data. It is done by a non-linear function called activation function. Multilayer perceptron uses Back-propagation algorithm [7] for supervised learning. We will discuss Back-propagation algorithm and activation function in brief.

3.4.1 Back-propagation algorithm

The back-propagation learning algorithm [7] has mainly two steps: i) propagation ii) weight update.

Propagation is of two types: a) Forward propagation b) backward propagation. In forward propagation function signals flow from input layer to hidden and output layers. Forward propagation of training input is used to generate the propagation output activations. In backward propagation error signals propagate from output to hidden and input layers. Backward propagation is used to generate details of all output and hidden neurons in input layer. It is used as feedback in input layers.

Weight update is done by the following steps a) multiply local gradient and input signal of neuron b) subtract a portion of the gradient from the weight. It can be expressed by the following formula

$$(Weight\ correction) = (learning\ rate\ parameter) \times (local\ gradient) \times (input\ signal\ of\ neuron) \tag{3.22}$$

3.4.2 Activation function

Hidden neurons are used to recover and extract features from the training data. It is done by a non-linear function called activation function [7]. Computation of local gradient is related to the derivative of the activation function. For that purpose, activation function is very important in Multi-layer perceptron. There are mainly two types of activation functions: a) Logistic Function: b) Hyperbolic tangent function

Logistic function: This type of activation functions are expressed by the following equation

$$\varphi_j(v_j(n)) = \frac{1}{1+\exp(-av_j(n))} \tag{3.23}$$

Where, $v_j(n)$ is the local field of neuron j and a is an adjustable positive parameter. Differentiating equation (4.2) with respect to $v_j(n)$, we get

$$\varphi_j(v_j(n))' = \frac{a\exp(-av_j(n))}{[1+\exp(-av_j(n))]^2} \tag{3.24}$$

Now from the activation function we can get local gradient as given by the following formula

$$\partial_j(n) = e_j(n)\varphi_j(v_j(n))' \tag{3.25}$$

3.4.3 Hyperbolic tangent function:

This type of activation functions are expressed by the following equation

$$\varphi_j(v_j(n)) = a\tanh(bv_j(n)) \tag{3.26}$$

$$\varphi_j(v_j(n))' = ab\sec h^2(bv_j(n)) \tag{3.27}$$

Now from the activation function we can get local gradient as given by the following formula

$$\partial_j(n) = e_j(n)\varphi_j(v_j(n))' \tag{3.28}$$

In our proposed Artificial Neural Network classifier we have used tan sigmoid activation function.

3.5 PERFORMANCE MATRICES

We have used four performance matrices [8] to analyse the classification performance of our proposed method in user independent condition. These performance matrices are : Accuracy (A_c), Sensitivity (S_e), Specificity (S_p) and Positive Predictivity (P_p). All of these parameters are given by the following equations [3]:

$$A_c = \frac{TP+TN}{TP+TN+FP+FN} \tag{3.29}$$

$$S_e = \frac{TP}{TP + FN} \tag{3.30}$$

$$S_p = \frac{TN}{TN + FP} \tag{3.31}$$

$$P_p = \frac{TP}{TP + FP} \tag{3.32}$$

In this equation *TP, TN, FN, FP* indicate true positive, true negative, false positive, false negative respectively. True positives are images which have been correctly assigned to a certain class whereas false positives are images which have been incorrectly assigned to that same class. A false negative occurs when an image should have been assigned to that class but was missed and assigned to another class. Similarly false positive is just opposite of false negative. Consequently, sensitivity measures how successfully a classifier recognizes images of a certain class without missing them whereas positive predictivity measures how exclusively it classifies images of a certain type.

3.6 RESULTS AND DISCUSSIONS

Moment features are calculated on the pixel values of images. For that reason, we have resized and cropped our original images into a (40× 40) size so that the hand region becomes the maximum region. We have empirically selected order of geometric moment as 49 (n=7 and m=7), Krawtchouk and Tchebichef moment as 64 (m=8, n=8).

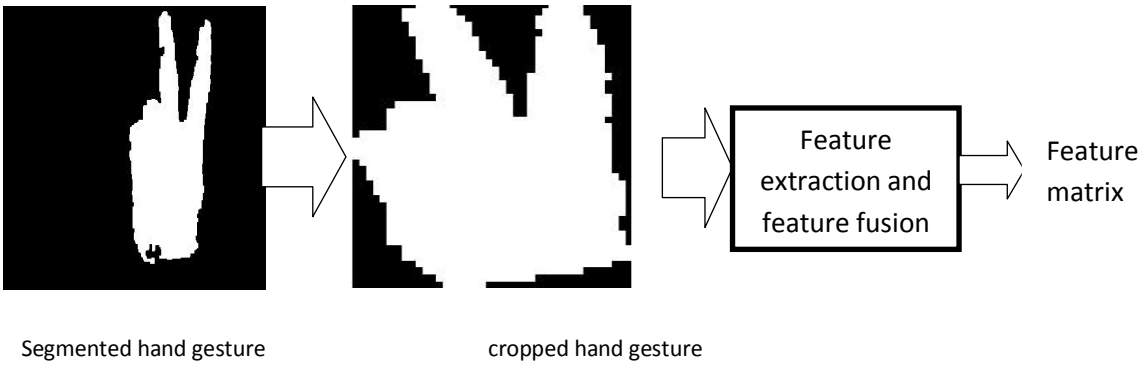


Figure 3. 2 Block Diagram

The dataset consists of 1500 color images of 10 gestures, 15 sample each class of 10 users. The dataset is equally divided into training and testing dataset of 750 images for 5 different users both to make the system user-independent. We have calculated our features differently for training and testing. For that reason, our training and testing dataset give 750 different features. As the order of Tchebichef and Krawtchouk moments are 64, these features consist of 64 values and feature size becomes 64. We have empirically selected translation parameter p as 0.5 to make the feature scale invariant. On the other hand, Geometric moment has an order of 49, so the size of feature is 49. So we got two feature matrices of size 750× 64 for Krawtchouk and Tchebichef moments and of size 750× 49 for Geometric moment. These training and testing feature matrices have been used in classification part. First the network is trained by training feature matrices and then tested using testing feature matrices.

Krawtchouk moment consists of a normalization factor. For that reason, we got all the values of feature sets within the range of -1 to +1. Further these feature values are divided by its maximum values to make the feature range [0, 1]. As all the values were in the range [-1, 1], after normalization, range of maximum and minimum values are quite appreciable. On the other hand, Geometric and Tchebichef moments don't have any normalization factor. So, the range of maximum and minimum values is not satisfactory after normalization.

To improve classification performance we have proposed two feature fusion strategies: Serial feature fusion and parallel feature fusion. In case of serial feature fusion, size of fusion feature is same as the original two features. In case of parallel feature fusion, size of the fusion feature is sum of the two features. So in serial feature fusion of Krawtchouk and Tchebichef moment feature size is 128. In serial fusion of Krawtchouk and Geometric moment feature size is (64+49) = 113. In case of parallel fusion of both Krawtchouk-Tchebichef and Krawtchouk-Geometric moment feature size is 64.

We have quantified our classifier performance using 4 matrices [8]: accuracy, sensitivity, Specificity and positive predictivity. Performance of three moments in terms of these four parameters is shown in table 1 and Fig.5. It shows that Krawtchouk moment is the best in terms of all the performance matrices. In user independent condition, neither of these moments has shown satisfactory classification accuracy or sensitivity. Geometric moment shows worst performance in terms of all these performance matrices. Krawtchouk moment, Tchebichef moment and Geometric moment have shown 91.53%, 82.67% and 76.2% classification accuracy respectively, as shown in table1 and Fig.5.

To improve classification accuracy we have implemented two feature fusion strategies as discussed in chapter 3. In that case, classification performance has improved significantly. For features with same size (Krawtchouk and Tchebichef moment) parallel feature fusion has given best result and for unequal size features (Krawtchouk and Geometric moment) serial feature fusion has given best result. Serial fusion of Krawtchouk-Tchebichef moment and Krawtchouk-Geometric moment give 93.53% and 94.93% classification accuracy as shown in table.1 and Fig.6. Parallel fusion of Krawtchouk-Tchebichef and Krawtchouk-Geometric moment give 95.33% and 94.2% classification accuracy respectively as shown in table 1 and Fig.6. It is clear that for both these two fusion strategy, classification accuracy has increased significantly. From table 1 it is clear that, for equal size features parallel fusion gives the best result and for unequal size features serial fusion gives the best result. In user-independent situation, serial fusion of Krawtchouk and Tchebichef moment has shown best classification performance in terms of all these matrices.

Features	Accuracy	Sensitivity	Positive Predictivity	Specificity
Geometric moment	95.24	76.2	76.06	97.36
Krawtchouk moment	98.31	91.53	91.87	99.06
Tchebichef moment	96.53	82.67	82.22	98.07
Geometric krawtchouk moment serial fusion	98.99	94.93	95.05	99.44
Geometric krawtchouk parallel fusion	98.84	94.2	94.32	99.36
Tchebichef krawtchouk serial fusion	98.71	93.53	93.85	99.28
Tchebichef krawtchouk parallel fusion	99.07	95.33	95.42	99.48

Table 3. 1 Performance comparison of various features

Class	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0	150	0	0	0	0	0	0	0	0	0
1	0	148	0	2	0	0	0	0	0	0
2	0	0	150	0	0	0	0	0	0	0
3	3	2	0	142	0	3	0	0	0	0
4	10	0	0	0	112	1	10	12	5	0
5	20	0	20	0	0	93	7	2	8	0
6	6	0	10	0	3	0	124	0	6	1
7	1	0	7	0	10	14	0	79	7	32
8	0	0	6	0	6	8	21	42	45	22
9	1	0	5	0	0	2	15	27	0	100

Table 3. 2 Confusion matrix of Geometric moment feature

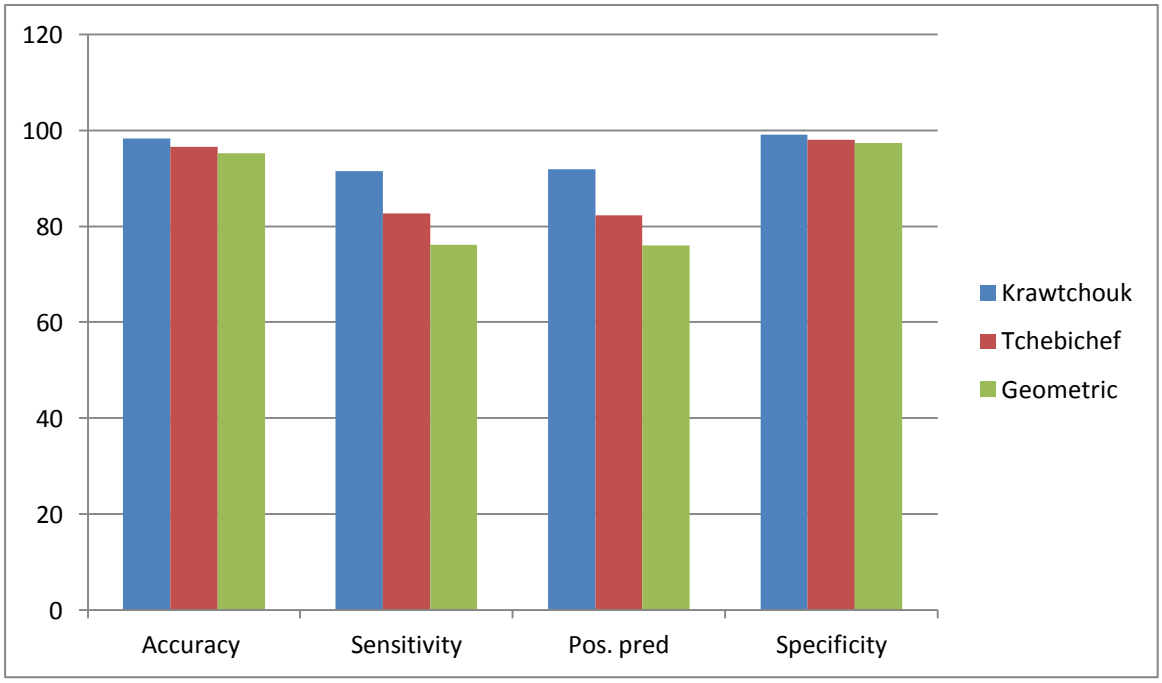


Figure 3. 3 Performance comparison of three moment features

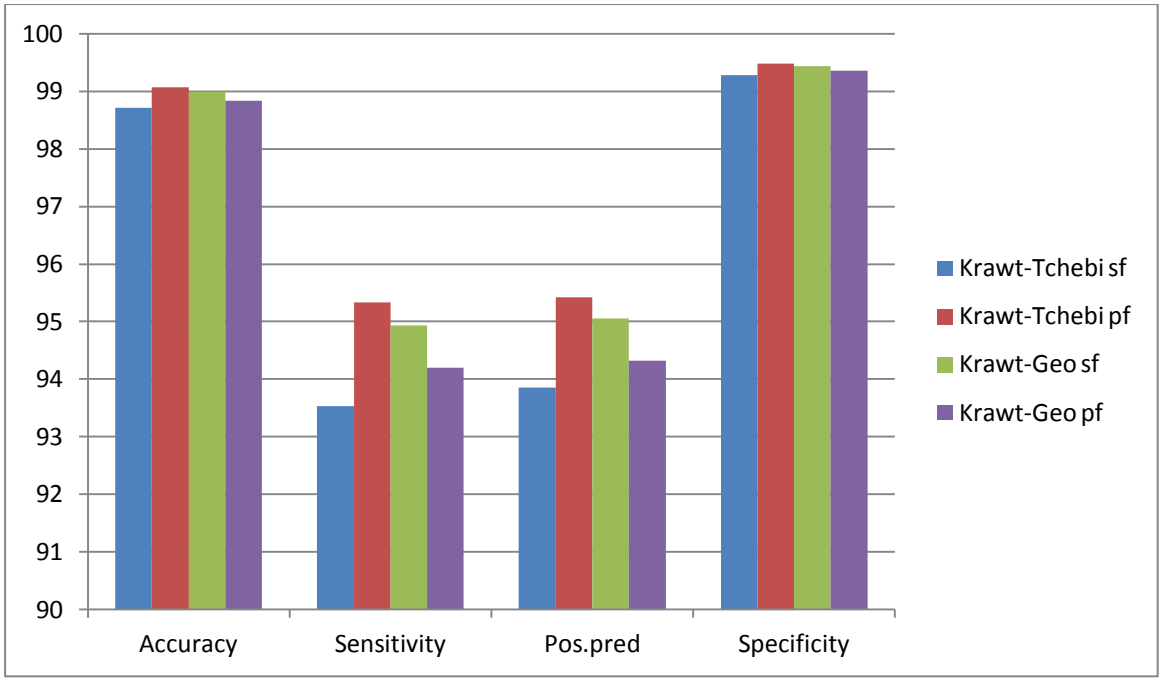


Figure 3. 4 Performance comparison of fusion features of moments

In user-independent gesture recognition, misclassification has occurred for geometrically closed gestures**.** In case of Geometric and Tchebichef moments, mismatches occur more than Krawtchouk moment as shown in Table.2, Table.3 and Table.4. This is because geometric moment is a local feature and it only represents the statistical attributes of the shape. On the other hand, although Tchebichef moment is orthogonal, it doesn't show satisfactory classification performance in user-independent condition. In case of Geometric moment, gesture 7 is misclassified as 8 and 9, gesture 8 is misclassified as 6, 7 and 9, gesture 9 is misclassified as gesture 6 and 7 as shown in table.2. In case of Tchebichef moment gesture 8 has been misclassified as gesture 7 and 9, gesture 9 has been misclassified as gesture 7 and 8 as shown in table.3. In case of Krawtchouk moment, mismatch occur less than Geometric and Tchebichef moments. In that case, gesture 9 has been misclassified as gesture 8, gesture 8 has been misclassified as gesture 7 and gesture 7 has been misclassified as gesture 9 as shown in table.4. In all these three moments most of the misclassification have occurred in case of gesture 7, 8 and 9 as these gestures have geometrically closed shapes.

To overcome this mismatch problem, two feature fusion strategies (serial fusion and parallel fusion) have been proposed. Experimental results show that fusion strategies significantly improve the gesture recognition performance and parallel feature fusion of Krawtchouk-

Tchebichef moment has given the best gesture recognition performance. In case of all these four fusion features, misclassification rate has been decreased significantly as shown in Table.4, Table.5, Table.6 and Table.7.

Class	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
0	131	0	6	2	0	0	2	0	1	8
1	0	150	0	0	0	0	0	0	0	0
2	0	1	147	0	0	0	0	0	1	1
3	2	0	0	145	0	3	0	0	0	0
4	0	0	1	0	120	0	15	14	0	0
5	0	0	0	20	0	124	0	0	1	5
6	0	0	1	0	10	0	136	0	3	0
7	0	9	6	0	3	0	12	106	10	4
8	0	2	15	0	23	0	7	30	58	15
9	7	0	1	0	0	1	1	2	15	123

Table 3. 3 Confusion matrix of Tchebichef moment

Class	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
0	138	0	10	0	1	1	0	0	0	0
1	0	150	0	0	0	0	0	0	0	0
2	0	0	150	0	0	0	0	0	0	0
3	0	0	0	132	0	1	4	0	0	13
4	0	0	0	0	150	0	0	0	0	0
5	14	1	0	6	12	112	0	0	0	5
6	0	0	0	1	0	0	148	0	1	0
7	0	0	0	0	0	0	0	132	1	17
8	0	0	1	0	1	0	1	10	133	4
9	0	0	0	0	4	1	4	0	13	128

Table 3. 4 Confusion matrix of Krawtchouk moment

Class	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
0	144	0	6	0	0	0	0	0	0	0
1	0	150	0	0	0	0	0	0	0	0
2	0	0	150	0	0	0	0	0	0	0
3	0	0	0	147	0	1	2	0	0	0
4	0	0	0	0	150	0	0	0	0	0
5	10	0	0	4	4	131	0	0	0	1
6	0	0	0	0	0	0	148	0	2	0
7	0	0	0	0	0	0	6	131	3	10
8	0	0	0	1	2	0	5	5	131	6
9	0	0	0	0	1	1	3	0	3	142

Table 3. 5 Confusion matrix of serial fusion of Krawtchouk and Geometric moment

Class	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
0	137	0	12	0	0	1	0	0	0	0
1	0	150	0	0	0	0	0	0	0	0
2	0	0	150	0	0	0	0	0	0	0
3	0	0	0	147	0	1	2	0	0	0
4	0	0	0	0	150	0	0	0	0	0
5	10	0	0	7	3	128	0	1	0	1
6	0	0	0	0	0	0	150	0	0	0
7	0	0	0	0	0	0	2	140	0	8
8	0	0	0	1	0	0	11	7	124	7
9	0	0	0	0	0	1	4	0	8	137

Table 3. 6 Confusion matrix of parallel fusion of Krawtchouk and Geometric moment

Class	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
0	146	0	1	1	0	2	0	0	0	0
1	0	150	0	0	0	0	0	0	0	0
2	0	0	150	0	0	0	0	0	0	0
3	0	0	0	146	0	1	3	0	0	0
4	0	0	0	0	150	0	0	0	0	0
5	11	0	0	5	6	119	0	0	0	9
6	0	0	0	0	0	0	149	0	1	0
7	0	12	1	0	0	0	3	127	2	5
8	0	8	1	1	0	0	11	0	124	5
9	0	0	0	0	0	1	3	3	1	142

Table 3. 7 Confusion matrix of serial fusion of Krawtchouk and Tchebichef moment

Class	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***
0	144	0	0	3	0	3	0	0	0	0
1	0	150	0	0	0	0	0	0	0	0
2	0	0	150	0	0	0	0	0	0	0
3	1	0	0	146	0	1	2	0	0	0
4	0	0	0	0	150	0	0	0	0	0
5	3	0	0	8	0	139	0	0	0	0
6	0	0	0	0	0	0	145	0	5	0
7	0	0	0	0	0	0	4	138	0	8
8	0	0	1	1	0	0	10	7	126	5
9	0	0	0	0	0	0	4	0	4	142

Table 3. 8 Confusion matrix of parallel fusion of Krawtchouk and Tchebichef moment

3.7 CONCLUSION

A novel feature fusion technique for static hand gesture recognition is proposed in this work, which overcomes the challenges of misclassification of geometrically closed gestures. In case of Geometric and Tchebichef moments, mismatches occur more than Krawtchouk moment. This is because geometric moment is a local feature and it only represents the statistical attributes of the shape. On the other hand, although Tchebichef moment is orthogonal, it doesn't show satisfactory result in user-independent condition. To overcome this mismatch problem, two feature fusion strategies (serial fusion and parallel fusion) have been proposed. Experimental results show that fusion strategies significantly improve the gesture recognition performance and parallel feature fusion of Krawtchouk-Tchebichef moment has given the best gesture recognition performance.

REFERENCES

1. S. P. Priyal and P. K. Bora "A Robust Static Hand Gesture Recognition System Using Geometry based Normalization and Krawtchouk Moments" *Pattern Recognition, vol. 46, no. 8, pp. 2202-2219,* 2013.
2. S. Gupta, J. Jaafar, and W. F. W. Ahmad, "Static Hand Gesture Recognition Using Local Gabor Filter". *In: International Symposium on Robotics and Intelligent Sensors (IRIS 2012), vol. 41, pp. 827-832,* Kuching, Sarawak, Malaysia (2012).
3. S. P. Priyal and P. K. Bora "A Study Of Static Hand Gesture Recognition using Moments" *International Conference on Signal Processing and Communications (SPCOM), pp. 1-5,* IISC, Bangalore (2010).
4. D. K. Ghosh and S. Ari "A Static Hand Gesture Recognition Algorithm Using K-Mean Based Radial Basis Function Neural Network" *In: 8th International Conference on Information, Communications and Signal Processing (ICICS),* pp. 1-5, Singapore , 2011.

5. T.William, T. Freeman and M. Roth "Orientation histogram for hand gesture recognition" *in proceedings of the 1st International workshop on Automatic face and gesture recognition, pp. 296- 301*, 1995.
6. J. Yang, Yang, D. Jhang and J. F. Lu "Feature Fusion: Parallel Strategy vs. Serial Strategy*". Pattern Recognition, vol. 36, no. 6 , pp. 1369-1381*, 2003.
7. Haykin, S.: Neural networks. *Prentice-Hall (1999, 2nd edn.).*
8. A. Daamouche, L. Hamami, Alajlan, N. and Melgani, F. "A Wavelet Optimization Approach for ECG Signal Classification". *Biomedical Signal Processing and Control, vol. 7, no. 4,pp. 342-349*, 2012.
9. Z. Zou, P. Premaratne, R. Monaragala, N. Bandara and M. Premaratne. "*Dynamic Hand Gesture Recognition System using Moment Invariants" 5th International Conference on Information and Automation for Sustainability (ICIAFs)*, Dec. 2010, Colombo, Sri Lanka.
10. S. Kong, X. Wang, D. Wang, and F. Wu. "Multiple feature fusion for face recognition" *10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Apr. 2013, Shanghai, China.
11. S. Q. Sun, G. S Zeng, Y. Liu, A. P. Heng. and S. D. Xia. "A new method of feature fusion and its application in image recognition" *Pattern Recognition vol. 38, pp. 2437-2448,* 2005
12. Y. Fu, L.Cao, G. Guo. and Huang. S. T. "Multiple feature fusion by subspace learning" *international conference on Content-based image and video retrieval (CIVR 2008), pp. 127-234*, New York, 2008.

CHAPTER 4

CONCLUSION AND FUTURE WORK

4.1 CONCLUSION

A novel feature fusion technique for static hand gesture recognition is proposed in this work which overcomes the challenges of misclassification of geometrically closed gestures. In case of Geometric and Tchebichef moments, misclassification of gestures occurs more than Krawtchouk moment. This is because geometric moment is a local feature and it only represents the statistical attributes of the shape. On the other hand, although Tchebichef moment is orthogonal, it doesn't show satisfactory result in user-independent condition. To overcome this mismatch problem, two feature fusion strategies (serial fusion and parallel fusion) have been proposed. Experimental results show that fusion strategies significantly improve the gesture recognition performance and parallel feature fusion of Krawtchouk-Tchebichef moment has given the best gesture recognition performance.

We have proposed a skin colour segmentation method based on Modified K-means clustering and Mahalanobish distance. All the previously used techniques for skin colour segmentation depend on pre-defined threshold values. With the variation in skin complexion and illumination level, this threshold values also vary. In our proposed method, hand region is segmented by a semi-supervised learning method based on Modified K-means clustering and Mahalanobish distance. This method has shown robustness in illumination and skin colour complexion variation.

To make our proposed static hand gesture recognition system real time efficient we have proposed an illumination normalization technique. We also have made our system user independent by using different users for training and testing purpose.

Parallel feature fusion of Krawtchouk and Tchebichef moment has shown the best classification performance in user-independent condition in terms of all the performance matrices. We got 95.33% classification accuracy in case of parallel feature fusion of Krawtchouk and Tchebichef moment.

4.2 FUTURE WORK

Hand gesture recognition systems can be used in many real time applications like human computer interaction (HCI), robot control, remote control, Sign language recognition etc. We have proposed a real time efficient and user independent static gesture recognition system. This work can be efficiently employed to any real time applications like VLC media player control, mouse control, remote control of a television system, robot control etc. using hand gestures.

We have recognized digit 0-9 of American Sign Language using our proposed method. Same methodology can be employed for recognizing American Sign Language alphabets (A-Z) and words.

We have made our feature fusion strategies based on fundamental set theory concepts. In case of more complex recognition problems, this feature fusion process may not show satisfactory classification accuracy. To improve and modify feature fusion algorithm we have to include some statistical independence concept in this work. Maximum Likelihood Estimator (MLE) and Canonical Correlation analysis (CCA) can be employed in feature fusion process.

We made some restrictions in our database like hand region has maximum area with respect to other objects in the image, user's face is not allowed in the database and forearm region is wrapped by a black cloth. To make our algorithm real time efficient, these restrictions must not be imposed.

PUBLICATIONS

S. Chatterjee, D. K. Ghosh and S. Ari "Static Hand Gesture recognition based on Fusion of Moments" *In 1st International Conference on Intelligent Computing, Communication and Devices (ICCD 2014)*, Bhubaneswar, Odisha, Apr. 2014.